Downloaded from orbit.dtu.dk on: May 05, 2023

DTU Library

# Coding and inferences from visual and other behavioural data

**Sønderstrup-Andersen, Hans Henrik Krogh; Andersen, Henning Boje; Hilburn, B.G.; Zon, R.; Blechko, A.; Ober, J.; Hauland, G.**

*Publication date:*
2004

*Document Version*
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

*Citation (APA):*
Sønderstrup-Andersen, H. H. K., Andersen, H. B., Hilburn, B. G., Zon, R., Blechko, A., Ober, J., & Hauland, G. (2004). *Coding and inferences from visual and other behavioural data*. Danmarks Tekniske Universitet (DTU). Denmark. Forskningscenter Risoe. Risoe-R No. 1474(EN)

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Risø-Report

RISØ

Risø-R-1474(EN)

# Coding and inferences from visual and other behavioural data

Hans H.K. Andersen[1], Henning B. Andersen[1], Brian G. Hilburn[2], Rolf Zon[2], Anastasia Blechko[2], Jan Ober[3], Gunnar Hauland[4]

[1]Risø National Laboratory, [2]NLR, [3]IBIB, [4]DNV

Risø National Laboratory
Roskilde
Denmark
July 2004

**Author:** Hans H.K. Andersen, Henning B. Andersen, Brian G. Hilburn, Rolf Zon, Anastasia Blechko, Jan Ober, Gunnar Hauland
**Title:** Coding and inferences from visual and other behavioural data
**Department:** Systems Analysis

**Risø-R-1474(EN)**
**July 2004**

**ISSN 0106-2840**
**ISBN 87-550-3360-1**

**Abstract (max. 2000 char.)**:

In this report we focus on developing and testing methods for the analysis of team situation awareness. That is, focus is on methods for analysing team situation awareness relating visual and other behavioural data; development of tools to facilitate usability and efficiency of Eye Point of Gaze (EPOG) data collection; and exploration and documentation of EPOG data measurement techniques.

The objective is to develop methods at a semantic level for analysing shared situational awareness relating visual and other behavioural data and to demonstrate how (a) objective measures (eye tracking) (b) subjective measures (Workload, Situation Awareness (SA), Self-assessment) and (c) elicitation of subjects' awareness of team-mates' current SA correlate. Measures of subjects' awareness of team-mates was developed and tested during small-scale experiments. Specifically focus has been the combination of visual and other behavioural data, of subjective (where raw data are interpretations) and objective data (where raw data are recordings of directly observable behaviour), and of data representing voluntary/intentional behaviour (actions) and EPOG data (that do not directly represent intentions).

A series of small-scale experiments was carried out in the following domains: Flight Simulation, Nuclear Power Plant Operation, Air Traffic Control Simulation and Anaesthesia Simulation.

A device has been developed (JAZZ) that provides operators with real-time as well as historic data on co-operators attention mode. The prototype is based on a model that distinguishes three modes of mental activity (described as planning, monitoring and exploration). According to this model, visual activity increases across the planning, monitoring and explorations modes respectively. It is the thought that this model could be utilised to give feedback to operators in terms of team-mates current mental mode and thereby support maintenance of team situation awareness.

**Contract no.:**

**Group's own reg. no.:**
(Føniks PSP-element)

**Sponsorship:**

**Cover :**

**Pages:**
**Tables:**
**References:**

Risø National Laboratory
Information Service Department
P.O.Box 49
DK-4000 Roskilde
Denmark
Telephone +45 46774004
bibl@risoe.dk
Fax +45 46774013
www.risoe.dk

# Contents

# 1 Introduction

WP2 of the VINTHEC II project consisted of 3 subtasks: Exploring methods for analysing shared situational awareness relating visual and other behavioural data: Development of tools to facilitate usability and efficiency of EPOG data collection; and exploration and documentation of EPOG data measurement techniques. Risø have together with NLR and IBIB focused on developing and validating novel methods for the analysis of shared situational awareness. This work is presented in this technical report. QinetiQ and BAE SYSTEMS have focused on usability and EPOG systems and exploration of data measurement techniques. The result of this work is presented in the VINTHEC II technical report 2-1 "EPOG Ergonomic Guidelines."
Please notice that in this document the terms "shared situational awareness" (sSA) and "team situation awareness" (TSA) both are used interchangeably. In the VINTHEC II project the partners have agreed upon using sSA. However, the authors of this document have decided to use the term TSA in those situations where they refer to (non VINTHEC related) work where the term TSA is more common.
This part of WP2 sought to develop methods at a semantic level for analysing shared situational awareness relating visual and other behavioural data. This WP demonstrates how (a) objective measures (eye tracking) (b) subjective measures (workload, SA self-assessment) and (c) elicitation of subjects' awareness of team-mates' current SA correlate. Measures of subjects' awareness of team-mates has been developed and tested via small scale experiments. Specifically, this part of WP2 has looked at the combination of visual and other behavioural data, of subjective (where raw data are interpretations) and objective data (where raw data are recordings of directly observable behaviour), and at data representing voluntary/intentional behaviour (actions) and EPOG data (that do not directly represent intentions).

IBIB, a partner to the VINTHEC II consortium, has developed a model of cognitive functioning that assures that different saccadic activities are directly related to certain types of cognitive states. The pilot's shared situational awareness develops through conscious attention activities. According to this model conscious attention can be involved in three different modes of mental activity: Exploration, monitoring, and planning. The exploration mode activities are predominantly based on sensor-motor operations during real-time control. During exploration, the majority of fixations are task relevant. Monitoring mode activities are based on previously learned skills in selecting the most appropriate routine for fulfilling a given task. Reactions are based on predictions of a dynamic model of the state of affairs in the task environment. A few or only one fixation on an area of interest is characteristic for this cognitive state (i.e. high demand situations). Different areas of interests are attended in what seems to be a chaotic and unrelated order.

The IBIB view also states that in planning mode no new visual input is needed, that is there is no demand for acquiring and evaluating the state of affairs in the operating environment. In this mode, the operator is exclusively occupied with the future or the past of system control. During this mode the eyes usually continue saccadic scanning, but less intensely. It is also possible that the saccadic activity nearly ceases, with nearly no fixation of over 500ms duration.

The IBIB prototype is built upon the idea that switching between mental modes can be recognised by a visual activity signal. The IBIB prototype is based on a model that distinguishes three modes of mental activity (described as planning, monitoring and exploration). According to this model, visual activity increases across the planning, monitoring and explorations modes respectively. It is the thought that this model could be utilised to give feedback to operators in terms of team-mates current mental mode and thereby support maintenance of shared situational awareness. IBIB has developed a device (JAZZ) that provides operators with real-time as well as historic data on crew members' attention mode. This might help eliminating certain actions that would otherwise disturb other operators' SA like untimely (distracting when in planning mode) verbal communication co-operator or prolonged planning mode without maintaining proper scanning of instruments. The device consists of a three-state indicator of other operators’ current attention mode, with the possibility to retrospecially display data of visual activity with planning, monitoring and exploration marked on the scale.

In other research behaviour was analysed in a “bottom up” way to derive an indication of meaning. That is, it is assumed that EPOG behaviour does not itself indicate inherent meaning, and that semantic analysis is required to achieve this. We have used visual dwells of task-relevant viewing areas as the unit of analysis along with fixation number on areas of interest, fixation duration, fixation transitions, pupil diameter, blink rate, and blink duration

The possibility to analyse shared Situational Awareness (sSA) depends on how data is collected. Archival data (from VINTHEC I) are of limited value in this regard, since they were collected without regard to the possibility of new analysis techniques and methods (e.g., eye track video) envisioned for VINTHEC II . Risø and NLR have conducted a series of small pilot experiments (Flight Simulation, Nuclear Power Plant Operation, Air Traffic Control Simulation and Anaesthesiology Simulation) in order to obtain the format of data needed for developing the proposed methods and to arrange the tasks around teamwork.

The major outcome of this part of WP2 is the prototype VINTHEC II methodology, later to be implemented in the main experiment WP7. WP2 has an important link to scenario definitions, i.e. to make sure that the relevant crew SA behaviour is present during the main experiment in WP7. Thus, WP2 personnel must work closely with subject matter experts to design the WP4-scenarios.

With respect to data collection in WP7, we suggest to use two head- and eye-tracking systems for measuring subjects' eye movements. In addition, we should audio- and video tape the teams to record their verbal and non-verbal communication activities. After each session we should carry out debriefing semi-structured interviews with each single subject using the Assessment of Team Situation Awareness
(ATSA) form that focus on team situation awareness questions, e.g., to which degree did you feel that your team-mate was aware of your activities and intentions and to which did you feel that he was uncertain of what you were doing?

# 2 Shared situational Awareness as Mutual Knowledge

This chapter discusses Shared situational Awareness (SSA) as a phenomenon involving people and artefacts situated in social organisation of work. The concept of (SSA) has recently attracted much attention. It has its roots in research on Situation Awareness (SA). Many of the ideas discussed in this chapter originate from WP1. But we also want to look at what we can learn from other research communities not traditionally related to research on SA. In doing so we will mainly look at research within the area of Computer Supported Cooperative Work (CSCW). We will focus on three main topics: distributed and shared knowledge (not CSCW related), workspace awareness, and social modes of interaction.

## 2.1 Distributed Knowledge and Shared Knowledge

Recent ethnographic research suggests that cognition can be described as an interaction between an actor and a physical and social situation. Suchman (1987) in a series of classic investigations observed ordinary people engaged in everyday problem solving. She observed operators trying to repair malfunctioning photocopying machines and concluded that cognition is situated within a social organisation of work practices.
Along side, this conclusion Hutchins (1990), based on his study of team navigation, found that cognition does not solely reside in the mind of an individual. It is distributed and shared among team members in authentic situations. Distributed cognition is an endeavour that seeks to understand intelligence at a systems level (Flor & Hutchins, 1991). It purports to do this by studying:

The representation of knowledge both inside the heads of individuals and in the world.
The propagation of knowledge among different individuals and artefacts.

The transformations which external structures undergo when operated on by individuals and artefacts.

The external representations, which are most closely examined, are goal required external structures created by actors in the system and those structures used to support these goal relevant structures. Such structures can include drawings, typed or written statements, utterances, gestures, and physical models.
Hutchins (1990), described how knowledge in co-operative tasks traditionally is assumed to be:

"...partitioned among individuals in an exhaustive and mutually exclusive manner....At the other end of the knowledge distribution spectrum one can imagine a system in which everyone knows everything about the task." (Hutchins 1990, p. 212)

Hutchins rejects both views in arguing that an exhaustive and mutually exclusive knowledge distribution pattern is very vulnerable to system breakdown - if one team member fails to perform the whole system will fail, and a system where everybody knows everything will be very expensive to produce. Instead, Hutchins argues that substantial sharing between the actors where experts have task knowledge that subsumes that of novices taking part in the task. Dividing the task into co-ordinated parts permits the novices to contribute to the task. Hutchins brings forth that in many human systems that distribution of task knowledge is a result of a movement in the system with increasing expertise, with knowledge as the entry level most redundantly represented and knowledge at the expert level least redundantly represented. Let us take an example: Person A knows how to carry out a certain task, but since A has a lot interaction with B he also knows a bit about B's task. B on the other hand knows about A's job because he once had that job. Furthermore, B knows a good deal of the activities of C because they share a certain piece of equipment. C has had both A and B job in the so he knows everything about their tasks.

Cooke et al (2000) discusses a similar concept - Team Knowledge (TK). They define Team Knowledge as:

"The collection of tasks- and team-related knowledge held by team-mates and their collective understanding of the current situation. Team performance will be maximised to the extent that team knowledge accurate appropriately apportioned among members, and structured in such a way as to support compatible assessments of the task situation and development of effective strategies to cope with it." (Cooke et al, 2000)

The researchers view team knowledge as a specialisation of Team Cognition. Team Cognition involves a wide range of cognitive phenomena at the team level such as team decision-making, team situation awareness[1] and team perception. They distinguish two types of TK: The Team Mental Model and the Team Situation Model. The first type is conceptually close to distributed knowledge (Hutchins, 1990). The second type is conceptually close to that of team situation awareness ((TSA) Orasanu, 1990; Robertson and Endsley, 1997).
The Team Situation Model is described as fleeting and dynamic. It is acquired during task using a Team Mental Model and world cues. It is situation specific and its function is to interpret a given situation in a compatible way. The Team Situation Model is the team's collective dynamic understanding of a specific situation.
The Team Mental Model is described as long lasting and exists prior to the tasks. It acquired through training and experience. It has a variety of forms. Thus it can be declarative, procedural and strategic. It has a variety of contents in terms of knowledge of team member roles and responsibilities and knowledge of team members skills, knowledge, beliefs, abilities, etc. as well as tasks specific knowledge such as understanding of task procedures and typical task strategies. The Team Mental Model provides a collective for team members to draw upon when task situations develop.

## 2.2 CSCW

The term 'Computer Supported Co-operative Work' (CSCW) can be traced to a workshop held in 1984. The organisers were Irene Greif from MIT and Paul Cashmann from DEC. The workshop focused on the

1 They define TSA as a team's understanding of a complex and dynamic situation at any one point in time.

possibility of developing computer tools to support actors engaged in cooperative work (Greif, 1988). A number of prominent researchers from different research areas, e.g., office information systems, coordination technology, hypertext and computer conference systems, were invited to join the workshop (Bannon, 1993). This event was followed up by the first CSCW conference held 1986 in the US. Since then conferences have been held alternately in EU and US.

Typically the conference topics are organisational aspects related to the introduction of CSCW-applications in work settings, research into CSCW architectures, the role of ethnographic methods in CSCW systems design, the development of CSCW design methodologies, the development of CSCW hypermedia in supporting asynchronous and synchronous collaboration, and discursive topics related to the development of a conceptual framework for CSCW. Contribution has come from a wide range of different research disciplines, e.g., computer science, human factors, human computer interaction, participatory design, ethnomethodology, cognitive and social psychology, organisation theory, linguistics, etc.

CSCW has a different conceptual and methodological orientation than is commonly found with traditional Human Computer Interaction (HCI) research. HCI has primary focused on individual work situations while CSCW computers as possible mediators of cooperative work by supporting possibilities for cooperation through shared information spaces and by supporting coordination aspects of work (Bannon, 1989a; Bannon, 1989b; Bannon, 1989c).

This shift means that the analysis of work settings will focus on the interaction of different actors as they co-ordinate different tasks and utilise different tools. The ability to co-ordinate activities and the process of interpretation and perception it requires relies upon a social organisation build of a body of skills and practises which allows different actors to recognise what each other is doing and thereby generate appropriate behaviour. In this context, we might conceive of this organisation as a form of SSA where the actors develop and maintain an interrelated orientation towards a collection of tasks and activities.

CSCW applications have often in the literature been categorised according to a 2x2 time and space matrix introduced by Johansen (1988). According to this type of categorisation, CSCW applications can be conceived as enhancing real-time communication and collaboration or asynchronous interactions. Furthermore, the CSCW applications can be categorised as to whether they support actors engaged in face-to-face interactions or distributed in many locations.

*Table 1 An example of the 2x2 time and space matrix introduced by Johansen (1988) for categorising CSCW applications.*

| | Same time | Different times |
|---|---|---|
| **Same place** | face-to-face interaction | asynchronous interaction |
| **Different places** | synchronous distributed interaction | asynchronous distributed interaction |

Researchers often consider SSA as belonging to the "same place - same time" dimension, or "same time-different places" (e.g., pilot - air traffic controller. Also Andersen and Hauland (2000) has studied operator TSA during a simple control task in a nuclear reactor where the task can be categorised along this dimension). But do we also have to consider the other dimensions "same places - different times" or "different places - different times?"

Fuchs et al (1995) have introduced a similar model. This model focuses explicitly on modes of awareness. Support for the different modes of awareness is based on a semantic net that represents the working context in terms of objects, tools, actors and resources. According to the table, actors may get informed dynamically about events that happened currently or that have happened in the past in the surroundings of their actual position in the work environment. The visibility of events is bound to the actor's current work occupation. In this way the model should prevent information overload.

*Table 2 An example of the 2x2 matrix introduced by Fuchs et al (1995) for categorising awareness modes*

| | synchronous | asynchronous |
|---|---|---|
| **Coupled** | what is currently happening in the actual scope of work | what has changed in the actual scope of work since last access |
| **Uncoupled** | what happens currently anywhere else of importance | anything of interest happened recently somewhere else |

Synchronous awareness has to do with events, that are currently happening, whereas asynchronous awareness considers events that have occurred in the past. Orthogonal to this classification Fuchs et al (1995) distinguish according to the current interest of the actor between coupled and uncoupled awareness. Coupled awareness denotes the kind of overview that is closely related to the current occupation of the actor. Uncoupled awareness applies in situations where information about events needs to be provided independent of the actor's current focus of work. In the next section, we will take a closer look at a CSCW framework for collaborative awareness.

## 2.3 Workspace awareness

In this section, we will take a closer look at the framework of Workspace Awareness (WA) that Gutwin, and Greenberg (1998, 1999) developed within the CSCW research area. The framework takes research on SA as a point of departure for a conceptualisation of WA. The researchers view WA as a specialisation of SA, but focus explicitly on teamwork. They stress that in a collaborative situation peoples' SA must involve both the domain and the collaboration: "The SA that involves collaborating in shared workspace is what we call workspace awareness." (Gutwin and Greenberg, 1999, p6)

The goal of researchers is to develop a descriptive theory of awareness for the purpose of aiding CSCW design. They synthesise and organise existing research on awareness, and extend this work through a conceptual framework. They define WA as:

> "the up-to-the-moment understanding of another person’s interaction with a shared workspace. Workspace awareness involves knowledge about where someone is working, what they are doing, and what they are going to do next. This information is useful for many of the activities of collaboration—for coordinating action, managing coupling, talking about the task, anticipating others’ actions, and finding opportunities to assist one another." (Gutwin and Greenberg, 1999, p1)

In addition the they provide the boundaries for the definition:

"First, workspace awareness is awareness of people and how they interact with the workspace, rather than awareness of the workspace itself. Therefore, it does not explicitly involve knowledge of the artefacts on their own (although this knowledge is clearly essential to workspace awareness). Second, workspace awareness is limited to events happening in the workspace; it is therefore restricted to being ‘inside’ the temporal and physical bounds of the task that the group is carrying out. This means that workspace awareness differs from informal awareness of who is around and available for collaboration, and from awareness of cues and turns in verbal conversation, both of which have been studied previously in CSCW (e.g. Borning and Travers 1991; Dourish and Bly 1992; Greenberg 1996) and linguistics (e.g. Clark 1996; Goodwin 1981)." (Gutwin and Greenberg, 1999, p5)

According to Gutwin and Greenberg (1999) WA is both a product and a process. The product is the state of understanding about another actor's interaction with the workspace that allows people to interpret events. The process is the repetitive cycle of extracting information from the environment, integrating this information with existing knowledge, and using that knowledge to direct further perception. They argue that the maintenance of WA involves several cognitive activities including preattentive processing, attention allocation, perception, working memory management, comprehension and projection. This view is in line with the work of Endsley (1995) on situation awareness.

Based on Neisser's (1976) cognitive model of how awareness is maintained, their WA framework is organised around three issues: "what kinds of information people keep track of in shared workspaces, "how people gather workspace awareness information," and "how people use workspace awareness information in collaboration." Gutwin and Greenberg (1999) follow the human factors research that focus on awareness as knowledge created through interaction between an agent and its environment. They identify four basic characteristics that run through prior work on awareness (Adams et al 1995; Norman 1993; Endsley 1995):

> "Awareness is knowledge about the state of some environment, a setting bounded in time and space. For example, the environment might be the airspace that an air traffic controller is responsible for, and their knowledge might include aircraft headings, altitudes, and separation, and whether these factors imply a safe or unsafe situation.
>
> Environments change over time, so awareness is knowledge that must be maintained and kept up-to-date. Environments may change at different rates, but in all cases a person must continually gather new information and update what they already know.
>
> People interact with the environment, and the maintenance of awareness is accomplished through this interaction. People gather information from the environment through sensory perception, and actively explore their surroundings based on the information that they pick up.
>
> Awareness is almost always part of some other activity. That is, maintaining awareness is rarely the primary goal of the activity: the goal is to complete some task in the environment. For example, the air traffic controller's task is to move aircraft through a region efficiently and safely, and although awareness may affect success, it is not the primary intent." (Gutwin and Greenberg, 1999, p5).

Gutwin and Greenberg (1999) argue that shared workspace setting makes workspace awareness a specialised kind of situation awareness:

"when someone works alone in a workspace, their activities and their SA involve only the workspace and the domain task. In a collaborative situation, however, people must undertake another task, that of collaboration, and therefore their situation awareness must involve both the domain and the collaboration. The SA that involves collaborating in a shared workspace is what we call workspace awareness." (Gutwin and Greenberg, 1999, p6)

The first part of their conceptual framework is a list of elements. WA is made up of some combination of these elements

*Table 3 Elements of workspace awareness relating to real-time activity (Gutwin and Greenberg, 1998, p3)*

| Category | Element | Specific questions |
|---|---|---|
| Who | Presence | Is anyone in the workspace? |
| | Identity | Who is participating?<br>Who is that? |
| | Authorship | Who is doing that? |
| What | Action | What are they doing? |
| | Intention | What goal is that action part of? |
| | Artefact | What object are they working on? |
| Where | Location | Where are they working? |
| | Gaze | Where are they looking? |
| | View | Where can they see? |
| | Reach | Where can they reach? |

The researchers divide the elements in two groups: those that deal with what is happening with another person (e.g. amount of activity, nature of actions, changes, and expectations), and those that deal with where it is happening (location of focus, view extents, area of influence, or objects in use).

The second part of the framework consists of a list of WA mechanisms. In citing prior research (Segal 1994; Norman 1993; Dix et al 1993; Hutchins 1990) Gutwin and Greenberg suggest three main sources of workspace awareness information, and three corresponding mechanisms that people use to gather it. The prior research points out, that people obtain information that is produced by people's overt behaviour in the workspace, from workspace artefacts, and from conversations and gestures. The basic mechanisms through which people gather this workspace awareness information are:

- Consequential communication: the visible or audible signs of interaction with a workspace. Watching someone work provides clues about their actions.
- Feed through: the observable effects of someone's actions on the workspace's artefacts. Seeing an object move indicates that someone is moving it. Feedback from the environment or overall workspace caused by the indirect effects of someone's actions
- Intentional communication: explicit communication through speech or (the authors term) gesture, often employing deictic2 reference. Utterances, expressions, or actions that are not explicitly directed at others, but that are intentionally public.

These information-gathering mechanisms are very similar to the list of social modes of interaction discussed in the next section.

## 2.4 Analysis of Shared situational Awareness using observational data

In this section we will review sociological methods which might be adapted to provide measures of shared situational awareness. In particular we will focus of the framework for social modes of interaction and in

2 The practice of pointing or gesturing to indicate a noun used in conversation is called deictic reference.

detail present a case study to illustrate the use of the framework for analysing social interaction modalities that actors apply in maintaining awards of other's task activities.

A number of sociological studies using ethnographic methods have shown that work is embedded in a sphere of social patterns of non-formal interaction. No matter existing formal prescriptions of work, the actors are engaged in and depend on non-formal activities in carrying out their work (Wynn, 1979; Suchman, 1983; Suchman, 1987). In addition, as pointed out by Middleton, (1988), actors maintain consistent interpretations of the course, structure and contents of work tasks through such informal work activities. Moreover, in many settings the coordination of individual activities, though complex in nature, is managed through the rich variety of intuitive interaction modalities of everyday social life (Schmidt 1994).

In engaging in teamwork actors generally become mutual dependent. They cannot fulfil the tasks on their own, so they have to rely on the contribution of other actors applying their different capacities, competencies, strategies and perspectives. Given their interdependence they need, in some way, to articulate their individual activities in joining their efforts. The term "articulate" in this context comes from the work of Strauss (1985), and Gerson and Star (1986). In this sense articulation means to allocate, co-ordinate, schedule, interrelate, integrate, etc., individual activities according to the dimensions of who, where, when, how, what, etc. The articulation work can be considered a type of second order activities or overhead cost in terms of the use of resources or time. The actors engage in these overhead activities because they would not on an individual basis be able to accomplish a certain task.

Teamwork is constituted by the fact that multiple actors are interdependent in their work. In other words, they are working in the same "field of work", that is, they are transforming and controlling a conglomerate of mutually interacting objects and processes. Thus, all teamwork involves and, indeed, is based upon interaction through changing the state of a common field of work. What one actor - A - is doing is of import to B and C in doing their work. The other actors - C and B - may to some extent be able to infer what A is doing from the changing state of the field of work. However, while collaborating via changing the state of the field of work is basic to all teamwork, it is rarely adequate. In fact, articulation of teamwork involves and, indeed, requires a vast variety of social modes of interaction that are combined and meshed dynamically and seamlessly in accordance with the specific requirements of the unfolding work situation and the means of communication available. As we see it there are four main interaction categories or modes of interaction.

**Maintaining reciprocal awareness:** The team could be involved in synchronous activities, by monitoring colleges’ location in a room, and to monitor their activities. Moreover, they could be engaged in explicitly making their own activities publicly visible to teammates by thinking aloud, humming, etc.
**Directing attention:** Actors attract the attention of team-mates to focus on certain features or emerging problems in the field of work by, for example, to position certain items in certain ways, by pointing or nodding at particular items.
**Assigning tasks:** Actors could for example allocate a task by nodding at a work object or by stating a verbal request.
**Handing over responsibility** of processes in the field of work, for example, by passing on the work object in question, or the interface of a control mechanism.

These social modes of interaction are combined and meshed dynamically and fluently to meet the requirements of a specific situation. The different modes of interaction cannot be ordered in any simple kind of way but is possible to point at a limited number of prominent dimensions of the modes of interaction. Some examples:

**Unobtrusive versus obtrusive**, that is, some modes of interaction can be disruptive in nature in relation to a colleagues’ line of work, while others are very conspicuous and therefore permit colleagues to carry on working.

**Embedded versus symbolic**, that is, to embed cues in highlighting certain items belonging to the field of work by for example marking them versus using a symbolic representation of the cues, which through its abstract function offers a higher degree of freedom regarding the manipulation of the cues**.**
**Ephemeral versus persistent**, that is, shared situational awareness only appears during the course of work and then disappears without leaving any trail to track. It is for example not immediately possible to trace activities like monitoring co-workers activities or to make one's own activities publicly visible.

It is possible to articulate the individual activities by these rich interaction and communication modalities of everyday social life. This is evidenced by several studies of co-operative work (see for example Hughes et al, 1998; Harper et al, 1989; for the studies on Air Traffic Control. Heath and Luff 1991; Heath and Luff, 1992 on the studies on Line Control Rooms in the London Underground.)

To give an example here let us take the study of Line Control Rooms on the London Underground (Heath and Luff, 1991; Heath and Luff, 1992). This study shows how actors maintain fluent reciprocal awareness regarding other actors' activities. In doing so the actors monitor each other's activities by overhearing other actors' radio or telephone conversations. In addition, they attract attention to activities, which are less visible to others, for example, when working with timetables and logs, by reading or thinking aloud or even by humming, singing, feigning momentary illness etc.

The operators in the control room co-ordinate train traffic and movement of passengers on a particular line, in this case London's Bakerloo Line. The control room can house several staff, but concern here is with two main actors: the Line Controller who co-ordinates the day-to-day running of the railway and the Divisional Information Assistant (DIA) who, among other things, provides information to passengers and to Station Managers. Both operators are able to monitor the state of the Bakerloo line traffic on a real-time display, a `fixed line diagram', which runs the length of the room. In addition, a paper timetable specifies train numbers, times, and routes; crew allocations, shifts, and travel; vehicle storage and maintenance; etc. The Controller can contact train drivers via a radio system. The DIA, on the other hand, can monitor platforms via a closed circuit television (CCTV) and provide information to passengers via a Public Address system. In addition the DIA can establish contact with Station Managers by touch screen phone. Coordination of train traffic and passenger movement is a domain specific characteristic of rapid urban transport:

> "Unlike other forms of transport, rapid urban transport systems do not provide a timetable to the public. Instead, passengers organise their travel arrangements on the assumption that trains will pass through particular stations every few minutes. When such expectations are broken, or travellers are unable to change at certain stations, or have to leave a train because the line is blocked, then the DIA needs to provide information and advice. The nature of such announcements varies with the circumstances of, and reasons for, their production." (Heath and Luff, 1992, p. 74).

Because the two controllers have to co-ordinate the movements of trains and passengers speedily and with minimal discomfort to the public, the activities of the Controller and the DIA require extremely close coordination. Accordingly, the operators have developed "a subtle and complex body of practices for monitoring each other's conduct and coordinating a varied collection of tasks and activities" (Heath and Luff, 1992, p. 73). One element of this informal, implicit and yet systematic articulation of responsibilities and tasks is "an emergent and flexible division of labour which allows the personnel to lend support to the accomplishment of each others' tasks and activities and thereby manage difficulties and crises" (pp. 73 f.).

The operators of the Bakerloo Line need to be able to articulate their activities tacitly:

"It is relatively unusual for the Controller or the DIA to tell each other what tasks they are undertaking or explicitly to provide information concerning: the changes they have made to the service, the instructions they have provided to other personnel, or the announcements they have made to passengers. Indeed, given the demands on the Controller(s) and the DIA, especially when dealing with emergencies or difficulties. it would be impossible to abandon the tasks in which they were engaged explicitly to provide information to each other as to what they were doing and why. And yet it is essential that both Controller and DIA remain sensitive to each other's conduct, not only to allow them to co-ordinate specific tasks and activities, but also enable them to gather the appropriate information to grasp the details of the current operation of the service." (Heath and Luff, 1992, p. 74).

Heath and Luff (p. 75) provides a striking example of tacit development of reciprocal awareness:
...Controller calls Driver...
Controller: Control to the train at Charing Cross South Bound, do you receive?
...Controller switches monitor to the platform...
Controller: Control to the train at Charing Cross South Bound, do you receive?
Driver: Two Four O Charing Cross South Bound
Controller: Yeah, Two Four O. We've got a little bit of an interval behind you. Could you take a couple of minutes in the platform for me please?
Driver: Over
Controller: Thank you very much Two Four O.
DIA: "Hello and good afternoon Ladies an Gentlemen. Bakerloo Line Information... "

"The announcement emerges in the light of the DIA overhearing the Controller's conversation with the driver and assessing its implications for the expectations and experience of travellers using the service. He transforms the Controller's request into a relevant announcement by determining who the decision will effect and its consequences. In this case, this is particularly the passengers at Charing Cross whose train is delayed as a consequence of a problem emerging on the Southbound service. [...] The DIA does not wait until the completion of the Controller's call before preparing to take action. Indeed, in many cases, it is critical that announcements are delivered to passengers as Controllers are making adjustments to the service. In the case at hand, as the call is initiated, we find the DIA progressively monitoring its production and assessing the implications of the Controller's request for his own conduct. The technology, and in particular the fixed line diagram, provides resources through which the DIA can make sense of the Controller's actions and draw the necessary inferences. At the onset of the call he scans the fixed line diagram to search for an explanation, or provide an account for, why the Controller is contacting a driver and potentially intervening in the running of the service. By the Controller's second attempt to contact the driver, the DIA is moving into a position at the console where he will be able to reach the operating panel for the Public Address system and if necessary make an announcement. On the word `couple', at which point he can infer the potential delay that passengers might incur, he grabs the microphone and headset in preparation for the announcement. In consequence, even before the Controller's call to the driver is brought to completion, the DIA has set the Public Address system to speak to the passengers on a particular platform and is ready to deliver the announcement." (Heath and Luff, 1992, pp. 75 f.)

In the example given above, the DIA's very looking for evidence is motivated and driven by virtue of the Controller's attempt to call a driver:

> "Activities such as telephone conversations with personnel outside the room, tracking a particular train with the CCTV, or discussions with Line Management concerning the state of the service, are, at least in part, publicly visible within the local milieu and ordinarily the bits and pieces available can be used to draw the relevant inferences." (Heath and Luff, 1992, p. 79)

Having noticed the Controller's attempt to call a driver, the DIA scans the fixed line diagram in order to provide an account for the upcoming intervention. That is, the DIA is not only able to overhear the Controller and assume that they have mutual access to the same information displays, but is also able to discern, through ``peripherally monitoring the actions of his colleague", where the Controller might be looking and what he might have seen. "The various information displays, and their use by particular individuals, is publicly visible and can be used as a resource in determining courses of action and for the mutual coordination of conduct." (p. 76)

For the operators to make sense of what each other is doing, the activities of the other must be interpreted in relation to the state of the field of work. Thus, the formation of the reciprocal awareness requires access to (much of) the same evidence regarding the current state of the field of work (the movement of trains, passengers etc.): The fixed line diagram and the station monitors, provide an invaluable resource for the DIA in producing an account for his colleagues' interventions in the running of the service" (p. 76). In particular, the common availability of various sources of information in the Line Control Room allows the DIA to assume that the current problems in the operation of the service noticed by the Controller are similarly available to the himself if he scans the various displays.

> "The 'public' availability of the technology within the Control Room, whether it is a fixed line diagram, a CCTV screen, a screen-based line diagram or an information display, and the visibility of its use, provide critical resources in the collaboration between Controller and DIA. [...] More importantly perhaps, the DIA and Controller can use the common sources of information as a reliable means of accounting for a broad range of actions and tasks undertaken by the other. [...] Moreover, their use of the fixed line diagram and the surrounding monitors of the console is publicly visible, and can be used to determine a particular activity in which the DIA or Controller is engaged, or, [...] to display a potential problem which is emerging within the operation of the service. The mutual availability of the various information displays, and the visibility of their use, are important resources for making sense of the actions of a colleague and developing a co-ordinated response to a particular incident or problem." (Heath and Luff, 1992, p. 76)

Now, the formation of reciprocal awareness is not only the product of a - more or less - passive (visual and auditory) monitoring of what others are doing but involves the complementary proactive process of conveying cues of one's own activities and concerns. Thus, where activities (such as reading the timetable or entering the details of incidents on the various logs) are less visible, the details of the activity may not be readily available to a the others. Making such `less visible' activities accessible to colleagues may for example involve reading or thinking aloud, humming, and so forth. The London Underground case provides an excellent example of how one operator actively directs the attention of another to some particular feature of the state of the field of work in a way that is more direct and effective than merely marking certain objects but still unobtrusive and inconspicuous:

> "On occasions, it may be necessary for the Controller to draw the DIA's attention to particular events or activities, even as they emerge within the management of a certain task or problem. For example, as he is speaking to an operator or signalman, the Controller may laugh or produce an exclamation and thereby encourage the DIA to monitor the call more carefully. Or, as he turns to his timetable or glances at the fixed line diagram, the Controller will swear, feign momentary illness or even sing a couple of bars of a song to draw the DIA's attention to an emergent problem within the operation of the service. The various objects used by the Controller and DIA to gain a more explicit orientation from the other(s) towards a particular event or activity, are carefully designed to encourage a particular form of co-participation from a colleague, but rarely demand the other's attention. They allow the individual to continue with an activity in which they might be engaged, whilst simultaneously inviting them to carefully monitor a concurrent event." (Heath and Luff, 1992, p. 81)

Now, in spite of the enormous flexibility, efficiency, and effectiveness of these informal and implicit modes of interaction, the coordination of the myriad activities of the Bakerloo Line at large is far too complex, far too distributed in space and time, and involves far too many actors and specialties to be managed by means of these modes of interaction. These large-scale cooperative activities are basically managed by means of a timetable:

> "The Underground service is co-ordinated through a paper timetable which specifies: the number, running time and route of trains, crew allocation and shift arrangements, information concerning staff travel facilities, stock transfers, vehicle storage and maintenance etc. Each underground line has a particular timetable, though in some cases the timing of trains will be closely tied to the service on a related line. The timetable is not simply an abstract description of the operation of the service, but is used by various personnel including the Controller, DIA, Signalmen, Duty Crew Managers, to co-ordinate traffic flow and passenger movement. Both Controller and DIA use the timetable, in conjunction with their understanding of the current operation of the service, to determine the adequacy of the service and if necessary initiate remedial action. Indeed, a significant part of the responsibility of the Controller is to serve as a `guardian of the timetable' and even if he

is unable to shape the service according to its specific details, he should, as far as possible, attempt to achieve its underlying principle: a regular service of trains with relatively brief intervening gaps." (Heath and Luff, 1992, pp. 72 f.)

The timetable requires continuous management by the operators:

"The timetable is not only a resource for identifying difficulties within the operation of the service but also for their management. For example the Controller will make small adjustments to the running times of various trains to cure gaps which are emerging between a number of trains during the operation of the service. More severe problems such as absentees, vehicle breakdowns or the discovery of `suspect packages' on trains or platforms, which can lead to severe disruption of the service, are often successfully managed by reforming the service. These adjustments are marked in felt pen on the relevant cellophane coated pages of the timetable both by the Controller and the DIA, and communicated to Operators (Drivers), Signalmen, Duty Crew Managers and others when necessary." (Heath and Luff, 1992, p. 73)

"Perhaps the most critical activity within the Line Control Room [...], is rewriting the timetable; a process known as `reforming' the service. Almost all problems which arise in the operation of the service necessitate `reformations', where the Controller, actually within the developing course of an event, reschedules particular trains, their crews, and even their destination, so as to maintain, for the practical purposes at hand, a relatively even distribution of traffic along the line." (Heath and Luff, 1992, p. 79).

However, as opposed to changes to the state of the field of work as represented by the fixed line diagram or the platform monitors, changes made to the timetable are not immediately and automatically conveyed to the other operators. The distributed management of the timetable may therefore give rise to inconsistencies in the cooperative operation of the line. In this case, the Controller handles this by thinking aloud when his is making changes to the timetable:

"It is essential that both colleagues within the Line Control Room, and personnel outside such as Duty Crew Managers, drivers and even Station Managers, are aware of these changes. Otherwise, these staff will not only fail to enact a range of necessary tasks, but will misunderstand the state of the service and make the wrong decisions. Reforming the service however, is an extremely complex task, which is often undertaken during emergencies, and it is not unusual for the Controller to have little time explicitly to keep his relevant colleagues informed.

One solution to this potential difficulty is to render features of their individual reasoning and actions `publicly' visible by talking though the reformations whilst they are being accomplished. The Controllers talk aloud, but this talk is not specifically directed towards a colleague within the Control Room. Rather, by continuing to look at, and sketch changes on the timetable, whilst producing talk, which is often addressed to oneself, the Controller precludes establishing a `recipient' and the interactional consequences it would entail. Talking through the timetable, whilst rendering `private' activities `publicly' visible, avoids establishing mutual engagement with colleagues which would undermine the ongoing accomplishment of the task in question. Consider the following fragment in which the Controller finishes one reformation and then begins another.

*...Controller reads his timetable...*

Controller: It's ten seventeen to ( ) hhhhhhh (4.3)

Controller: Right (.) that's that one done.

Controller: hhh hhh (.) hhh

Controller: Two O Six ( ) Forty Six (0.7 )

Controller: Two Two Five

... the DIA begins to tap on his chair and the trainee begin a separate conversation. As they begin to talk the Controller ceases talking our loud...

Whilst looking at the timetable, the Controller announces the completion of one reformation and begins another. The Controller talks numbers, train numbers, and lists the various changes that he could make to the 206 to deal with the problems he is facing, namely reform the train to ~46 or to 225. As the Controller mentions the second possibility, the DIA begins to tap the side of his chair, and a moment or so later, discusses the current problems and their possible solutions with a trainee DIA who is sitting by the DIA's side. As soon as the DIA begins to tap his chair and display, perhaps, that he is no longer attentive to his colleague's actions, the Controller, whilst continuing to sketch possible changes on the timetable, ceases to talk out loud. Despite therefore, the Controller's apparent sole commitment to dealing with specific changes to the service, he is sensitive to the conduct of his colleague, designing the activity so that, at least initially, it is available to the DIA and then transforming the way the task is being accomplished so that it ceases to be `publicly' accessible. Whilst `self talk' may primarily be concerned with providing co-present colleagues with the necessary details of changes made by the Controller to the running order of the service, it is interesting to observe that a great deal more information is made available in this way than simply the actual reformations. [...]#[T]he Controller renders visible to his colleagues the course of reasoning involved in making particular changes. The natural history of a decision, the Controller's reasoning through various alternative courses of action, are rendered visible within the local milieu, and provides colleagues with the resources through which they can assess the grounds for and consequences of `this particular decision' in the light of possible alternatives. While the Controller is talking out loud, it is not unusual to find the DIA following the course of reasoning by looking at his own timetable, and where necessary sketching in the various changes which are made. In this way, DIA and Controller, and if present, trainees and reliefs, assemble the resources for comprehending and managing the service, and preserve a mutually compatible orientation to the `here and now', and the operation of the service on some particular day. The information provided through the various tools and technologies, including the CCTV monitors, the fixed line diagram, and information displays, is intelligible and reliable by virtue of this collaborative activity." (Heath and Luff, 1992, pp. 79-81)

In sum, then:

1. The field of work of the operators in the Bakerloo line control room, i.e., the trains and the infrastructure of the line on one hand and the passengers on the other, is not causally coupled in any strict sense Rather, the general function of the line operators is to establish a very close coupling of the movement of trains and passengers so as to provide the required quality of service to the passengers.
2. The various information displays, and their use by particular individuals, are publicly visible and can therefore be used as a resource in determining courses of action and for the mutual coordination of conduct. The operators can use the common sources of information as a reliable means of accounting for a broad range of actions and tasks undertaken by the other. The mutual availability of the various information displays, and the visibility of their use, are important resources for making sense of the actions of a colleague and developing a co-ordinated response to a particular incident or problem.
3. The operators do not regulate the state of the field of work by means of effectors or other control mechanisms. Rather, they regulate the state of the field of work by means of talking with train drivers, station managers, and passengers via radio and telephone. Accordingly, the two operators can develop and maintain a more rich and accurate reciprocal awareness by overhearing each other's conversations over telephone or radio.
4. The operators direct the attention of their colleague to certain features or events in myriad ways: by modulating their conversations with third parties, by humming or singing, by gazing etc.
5. The teamwork of the Bakerloo Line as a whole is basically managed by means of a timetable. To serve this coordination purpose, the timetable requires continual management by the operators.

This management of the timetable is itself a collaborative activity whose articulation may require the application of a whole repertoire of modes of interaction.

## 2.5 Shared situational Awareness

The notion of Situation Awareness has been used extensively since about the early 1980's in the human factors (HF) literature, and especially so in aviation HF writings. In this sub-chapter, we will provide a brief overview of Situation Awareness and we extend this notion to Shared situational Awareness, offering a definition of the latter. Finally, we outline some methods and techniques of measuring Shared situational Awareness. We describe a framework for describing 'Shared situational Awareness and exemplify the framework by reference to events contained in Air Traffic Management incident reports made by the Swedish Civil Aviation Authorities. The framework adopts as a central notion the concept of mutual knowledge, adapting this from the philosophy of language and theory of speech acts (Clark (1996) and others) and is inspired by recent work on Team Knowledge (Cooke et al., 2000 see above). The framework seeks to capture situations when team-mates fail to notice that their colleagues have a wrong or inadequate knowledge of current task features. Finally, we draw some conclusions for methods of measuring SSA failures - which might have no observable impact on team performance vis-à-vis the target system - and we recommend that more emphasis should be paid to eliciting estimates from team members of fellow team members' confidence in processing task variables.
SA was originally introduced in order to characterise – and, ultimately, to support the explanation of – how otherwise skilled human operators lose control of the system they are operating. So, the notion of SA was meant to be useful in explaining a certain type of control failure in the context of human-system interaction. The types of breakdown of system control, which the SA concept was meant to capture, involve:

- Human operators losing track of or missing task cues
- Human operators failing to integrate available cues into a coherent or valid mental model of the system-to-be-controlled

The motivation for appealing to SA in the context of human-system interaction is that there is a large cluster of control break-downs, which seem to share a number of features. It might therefore be useful to compare loss of situational awareness with a *syndrome* – namely, a cluster of symptoms of failures to maintain control. What these symptoms have in common is that they signify that the operator has lost grasp of dynamic features of the system-to-be-controlled. This does not mean that SA represents any "aggregate" cognitive function. Rather, SA is the dynamic state of an operator that relies on the proper functioning of a very broad range of cognitive skills.

We shall suggest the following starting points for the following review of the various definitions of SA and of the decomposing of the cognitive functions required for maintaining SA:

- It is customary within cognitive psychology to make a key distinction between individual abilities and the application of these abilities. Thus, the abilities or function underlying SA are different from the manifestation of SA in a given situation. It is primarily the manifestation of the SA phenomenon, and the measuring of this phenomenon, that is of interest here.
- SA is conceptually different from the variables that affect it. Thus, variables, which are claimed to indicate sufficient or insufficient SA, should not at the same time be postulated as constituting causes of sufficient or insufficient SA. Making this distinction may prevent circularity in the argumentation and it helps establishing a clear definition of the SA concept.
- SA is a psychological concept. This means that the term awareness labels a phenomenon that exists in the mind of individuals and that the term situation labels phenomena outside the mind of individuals. The traditional dichotomy between an inside mental world and the context is

important both for defining the concept and for developing measurement methods. Still, the usefulness of the SA concept depends on whether it aids in explaining how situation and awareness are related, i.e. the continuous interactions between operators, their systems and the environment. The manifestation of SA will be regarded as a product of many internal (cognitive) and external (situation) variables. It is assumed that this product has an objective existence in the operator's mind and that it can therefore be studied at the level of individuals. Thus, there is no built-in contradiction in using a cognitive information-processing framework for analysing SA, including team aspects.

SA has been described in cognitive terms, behavioural terms and in relation to specific system components – usually associated with the aircraft (Shrestha, Prince, Baker and Salas, 1995). Still, there is no uniformly accepted framework in which to define the SA concept, nor is there any commonly accepted operational definition of SA (Sarter and Woods, 1991; Weiner, 1990; Rogers, 1990, In: Shrestha, Prince, Baker and Salas, 1995).

The most widely known and cited definition of SA stems from Endsley, who, originally, targeted the term for use within the aviation domain (1988; 1990; 1995; see Endsley 1999 for a recent overview). While the concept Endsley describes is not new - she traces the SA concept back to studies of fighter pilots during World War I – she suggests a new theoretical and empirical framework for the further development of SA methodology. On Endsley's original definition – reiterated in her later papers - SA is *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future* (1988, p 97). Correspondingly, Endsley defines three levels of SA in direct parallel with the three stages of cognition outlined in the conceptual definition:

Level 1: Perception of environment (detection)
Level 2: Comprehension of current situation (identification)
Level 3: Projection of future status (prediction)

The two generic types of SA-breakdowns we described above correspond, thus, to Endsley's two first levels. A level 1 breakdown of SA involves the operator missing some cues in his environment. The kinds of cognitive functioning that may have failed when a level 1 SA breakdown takes place can be very varied - it may be prospective memory (forgetting to monitor) or low vigilance or overload due to excessive workload. Similarly, a level 2 breakdown of SA involves the operator failing to identify the current situation in terms of the generic types (e.g., engine failure, damaged undercarriage, potential conflict etc.) he has available in virtue of his professional training.

In more recent years, Bainbridge (1990) has discussed similar SA ideas, but used the phrase *mental picture*. A special issue of the journal Human Factors on SA appeared in 1995 - and this issue demonstrates that various authors define the SA concept differently. Still, many of the conceptual definitions seem to have some main ingredients in common.

Shrestha, Prince, Baker and Salas (1995) have reviewed SA definitions and identified five key situational attributes. The various definitions have been summarised under these headings, i.e. what to be aware of:

Awareness of the surroundings.

- Temporal awareness in dynamically changing situations.
- Awareness of mission objectives.
- Ability to observe, integrate, assess and act upon task relevant information.
- Anticipation of future events.

The following are some examples of conceptual SA definitions, emphasising the above criteria. Bolman (1979) referred to SA as the crew's *theory of the situation*. The crew must constantly challenge this theory so that erroneous theories can be discovered and corrected. The crew must communicate all relevant information and share the workload in the cockpit. Hollister (1988) described SA as the sum total of pilot's knowledge of the current situation and his/her role in it. Harwood, Barnett and Wickens (1988) referred to SA as the pilot's knowledge of a dynamically changing situation with respect to: spatial orientation, threats and system status, knowledge of whom is in command and mission progress. The temporal aspect is emphasised in this description of SA. The past is used to understand the present, and past/present is used to predict future events. Fracker (1988) defined situation awareness as the knowledge that results when attention is allocated to a zone of interest at a level of abstraction. The zone of interest refers to the concentric volumes of space surrounding the pilots (Endsley 1988). Level of abstraction refers to the methodological reduction, i.e. a selection of measures and resolution of the unit of analysis. Prince and Salas (1989, 1993) defined SA as a cluster of behaviours (through theoretical end empirical investigations), including perception of the surroundings and to the ability of identifying problems and recognising need for action. Kass, Herschler and Companion (1990) defined SA as skilled behaviour, namely the ability to extract, integrate, assess and act upon task relevant information. Thus, SA is linked to the cognitive knowledge-rule-skill based perspective of Rasmussen (1986). Kass et al. Described SA as skill based pattern recognition. Schwartz (1990) described SA as an accurate perception of the factors and conditions that affect an aircraft and its flight crew during a defined period. Sarter and Woods (1991) have referred to SA as the "accessibility of a comprehensive and coherent situation representation which is continuously being updated in accordance with the results of recurrent situation assessment".

It has been pointed out that any definition of SA should make it clear whether SA is either exclusively knowledge (state of awareness) or exclusively process (Smith and Hancock, 1995). Smith and Hancock suggest that SA *is* consciousness: First, SA behaviour must be directed towards an external goal. There is a distinction between directing the consciousness inwards, i.e. *introspection* versus directing it outward, *situation awareness*. Second, the environment (including the system to be controlled) determines what the agent must know and do.

### 2.5.1 A pragmatic definition of SA

Different SA definitions may place different emphases on cognitive aspects of SA, the situation aspect of SA and the cognition-in-context (interaction) aspect of SA. Within EU project DIVA (WP3.1 Evaluation Methodologies) a definition of SA has been proposed (by the authors of this subchapter) which will be adapted to this project. It was observed that in the in the case of the DIVA project - concerned with the development and HF testing of novel cockpit interfaces - it is desirable to establish a definition of SA which is pragmatically motivated, and which theoretically well founded and empirically supported. The same applies to VINTHEC II . As will be made clear in the following sections, this means that the framework we propose is intended to be *inclusive and comprehensive* with respect to the *aspects* of cognitive functioning which are suggested to underlie the upholding of SA. At the same time, the elements we suggest form part of SA are – in brief outline – the main components of broadly accepted information processing models of cognitive performance. We suggest that an adequate framework of SA should accommodate:

Building on the definition of SA as laid out in VINTHEC I (WPs 1 (VINTHEC, 1996) and 2) SSA as laid out in VINTHEC II WP1 can be generally described as knowledge of current and future state (VINTHEC II, 2001), which implicitly considers both the system goals and strategies required to meet these goals (VINTHEC II, 2001)

- The operator's awareness of the specific situational *goals* (possibly team goals), including the strategies to fulfil them.
- *Attention* towards task relevant elements in the situation, including possibly tasks and performance of colleagues

- Relevant, and thus appropriately updated, *knowledge* about system and environment parameters and states, including the *temporal* aspects of (continuous) changes.

When relevant, the mutual awareness between collaborating operators, say between a captain and a co-pilot and between the crew and other agents (ATC, cabin crew, other aircraft etc.)

### 2.5.2 Team aspects

So far team aspects have been alluded to only briefly. However, especially in the context of civil aviation it is necessary to expand the notion of individual SA to crew or team SA. However, it may be asked, is not team SA in the cockpit environment just the combined individual SA of the captain and of the co-pilot? Now, it may be useful and even required that the captain and the co-pilot do not have *divergent* representations of their a/c and the environment. But this is not enough. Each of them will need to have not only a representation of their system and environment, but also of what the other pilot knows and does not know.

Schematically, two persons, A and B, who are working together on a dynamic task, may have *coinciding awareness* with respect to a situation. This means that A and B perceive and comprehend the situation in the same way. But when A and B have *shared awareness* with respect to a situation or some parameter in their environment then A and B perceive and comprehend the situation in the same way; and, moreover, they *know that they share this awareness of the situation*. At the same time, A may not know the value of certain parameters but may know that B knows it; and vice versa. It has often been pointed out that s*hared* awareness of the current situation is the basis for efficient task allocation. Detailed studies and theories of the *recognition of others' beliefs and intentions* have been made within linguistics (pragmatics), not within cognitive psychology in a narrow sense. See e.g., Grice 1982; 1996).

A range of cognitive activities in the cockpit environment requires individual SA but not shared SA; conversely, some activities do require a robust shared SA. At the same time, incident and accident reports have demonstrated that failures to achieve or maintain shared SA – shared awareness – be dangerous in certain situations.
For some practical purposes it is often sufficient to measure crew situation awareness by measuring whether their awareness *coincides*. For instance, experimenters may question the subjects (independently) about the value of some task parameter and it may, for all practical purposes, be sufficient to record whether they give the same answer or not. If they give different answers, then their knowledge is not the same; but if they give the same answer, neither of them might possibly know what the other knows.

The following excerpt from Swedish CAA incident reports of ATC occurrences (loss of separation) illustrates breakdowns of Shared situational Awareness.

case 1. "He did not inform her, since he thought it was routine [and he therefore thought she knew]"

Following Salas et al (2000) we distinguish between an operator's mental model and his situation model. An operator's mental model is knowledge stored in long-term memory about his work domain and his tasks. Two ATC operators from different area control centers - let us say in the same country - will share a great deal of knowledge. But in addition, each of them will know a very large number of details about his or her sector or sectors - details which are so comprehensive that it takes three to six months to master them. This knowledge about his domain and his sectors is the operator's mental model. Then, when he comes to work and opens up his position and works for about an hour or so, the mental model is applied to - and serves to interpret the otherwise entirely chaotic cues of - the current situation. The operator's knowledge of the current situation is his situational model. The situational model is an "instantiation" of the domain mental model -, roughly in the way that my (generic) mental model of how a car works serves one when they start to drive a totally unknown model of car.

Now, a part of an operator's mental model of his task domain is his knowledge of what the standard operator is able to do and what he isn't, what he may be expected to do and what not. So an operator has a more or less permanent mental model of his task domain and a more or less permanent model of the "standard" operator (or standard colleague). A pilot in a large fleet will have professional expectancies as to the competencies of a fellow pilot in this fleet. If he happens to know already the other pilot, his picture of his colleague is even more detailed.

Similar to the distinction between the more or less permanent mental model and the current situational model, we may therefore distinguish between an operator's generic model of his colleagues and his situation determined model. This relationship has been modelled in computational linguistics (pragmatics) where one speaks of a speaker's interlocutor model - i.e., his model of other speakers in general (in these roles, perhaps - say, a salesperson in a shop) and of his current interlocutor in particular, and even of his current interlocutor here and now .

Shared situational awareness, we argue, may be distinguished in terms of the degrees to which it is valid to characterise it as mutual knowledge. At the one extreme, operators may communicate at an automatic or skill-based level without consciously thinking of the information needs of others. At the other extreme, an operator will carefully consider his colleagues' information needs and the resources she has currently available.
On our proposed definition of SSA a team (or pair) of operators have SSA to the extent they have mutual situational knowledge.

SSA - and hence, mutual situational knowledge involves knowledge of each other's current SA (and, by the definition of mutuality, it therefore also means that not only does A know roughly how B's current SA is (and vice versa, B knows A's) and very importantly, mutual knowledge about current priorities and current trust and mistrust (confidence and uncertainty).

### 2.5.3 Measures of communication correlated with performance outcomes

In this section,3 we review a number of studies that have examined the relation between crew communication and performance outcomes. These studies have examined communication patterns of crews either collected during flights (revenue operations) or during simulations involving high fidelity flight simulators (typically full flight simulators). Data collected during line operations have obvious face validity but there are obvious and nearly always-insurmountable obstacles in obtaining control over variables, so there are problems in generalising results. Conversely, simulations allow for control of most of the variables but the data they yield may not be entirely representative of real world operations.

While a few studies have looked as well at pre-flight communication patterns (pre-flight briefings), nearly all data about the relation between communication and crew performance are derived from observations and transcriptions of cockpit voice recordings, possibly augmented with video recordings of non-verbal communication. Studies of crew co-ordination and communication patterns were initiated in the late 70's, prompted in part by the observation that a number of accidents seemed to have involved a lack of efficient co-ordination between or among crew members.

The landmark study of crew co-ordination and communication is Ruffell-Smith's (1979) often-cited experimental simulation involving 18 airline crews flying a two-segment flight in a 747 simulator. The study was not originally designed an examination of the relations between performance and group and communication variables, but was targeted at studying the effects of stress and involved a technically complex scenario (hydraulics failure, bad weather, pestering cabin staff/VIP passenger, complex ATC

---

3 The section is based on a re-analysis of the work carried out in the DIVA project (Project funded by the European Commission under the Industrial and Materials technologies Programme (Brite Euram III) Contract number: BRPR-CT97-0551 Project number: BE96-4120).

instructions). As it turned out, the experiment did demonstrate great variation in performance and mission success among the crews. However, the major sources in variation in outcome were, not technical proficiency and piloting skills, but lack of crew co-ordination and inadequate communication. The cockpit voice data were subsequently analysed by Foushee and Manos (1981), who demonstrated a number of dependencies between 1) performance outcome and 2) co-ordination and communication variables (see below). In particular that the crews who exchanged more information about flight status committed fewer errors and experienced a better outcome than the others. A number of other differences among the technically similarly qualified crews were also observed: some captains allocated many tasks to first-officer and engineer, other captains largely failed to do so, engulfing themselves in the problems. Ruffell-Smith's original study and the follow-up analysis of Foushee and Manos introduced the basic techniques of communication measurements used in the following two decades: the voice recorder data were transcribed and the individual *speech acts* by each crew-member were *coded into pre-defined types.* That is, each utterance was grouped into one of small range of categories:

- Command
- Observation (about flight or system status)
- Inquiry or request for information
- Response uncertainty
- Agreement
- Acknowledgement
- Repetitions (of already stated commands or inquiries).

Most of the subsequent work, involving measures and analysis of communication, has involved variations on the techniques initiated by Ruffell-Smith and Foushee & Manos. Cockpit speech has been transcribed and coded into a fairly small number of speech acts categories. The analysis of relationships between communication patterns has consisted in exploring statistical correlations between patterns and frequencies of speech acts (by crew member or crew position) in relation to both outcomes and task or environment variables.

Foushee, Lauber, Baetge and Atcomb (1986) performed another influential and highly illustrative simulator study, again originally not targeted at communication processes as such but at examining the effects of fatigue. In this simulator study (Boeing 737-200) half the crews were rested, half were fatigued. The rested crews were in a state similar to when they report to work after two days rest and the fatigued crews had flown for two days before the target scenario, so they were comparable to crews at their final day of a three-day trip. Prior to the scenario session, the fatigued crews did report, as intended and expected, less sleep and more fatigue than the rested crews. However, the results of the study were not, as expected: the fatigued crews turned out to perform much better than the rested crews. The interpretation of this result was not, of course, that fatigue does not impact on safety and efficiency, but that the fatigued crewmembers had benefited from the fact that they had become familiar, within each crew, with one another. The benefit, which the fatigued crews derived from having just worked together for two days, was translated into a much more efficient crew co-ordination and this far outweighed the penalty imposed by fatigue.

In the following, we review findings about relations or lack of relation between performance outcomes and communication parameters.

#### 2.5.3.1 Amount of crew communication in relation to performance

Analysis of the communication data from the classical Ruffell-Smith experiment - as coded and analysed by Foushee and Manos (1981) - indicated that "crews who communicated more overall tended to perform better…" (Helmreich & Foushee, 1993). Indeed, this supposed relation is often invoked in CRM training programs. Nevertheless, a number of other studies have failed to duplicate the finding and, for instance, Wiener (1993) observed that "I am hesitant to conclude, pending further evidence, that quantity of communication is a hallmark of a good crew" (p. 213). Thus, as reported in Prince et al. 1997, Oser, Prince & Morgan (1990) used a modified version of the same categories of speech defined by Foushee and Manos (see bulleted list in previous section) studying 14 crews of military pilots flying a scenario in a full flight simulator. While they replicated Foushee and Manos' result that communication patterns vary across flight segments they found no differences in frequencies of communication relating to crew performance. Similarly, Andersen et al. (1996) used similar categories of communication to code bridge communication from 53 ship captains and crews (93 voyages; two scenarios) in a full mission marine simulator collecting more than 100 hours of bridge communication. The data from this study similarly revealed *no* correlation between performance and frequency of communication.

#### 2.5.3.2 Types of speech acts and performance

While results are mixed with respect to correlations between performance outcome and the combined or total amount of communication per session segment, it is natural to analyse the relation between types of speech acts (commands, inquiries, etc.) and performance outcome, possibly supplied with data about the position of the speaker and the hearer. In the Andersen et al. (1996) study of 100+ hours of ship bridge communication, analysis was made of the correlation between outcome and each of the speech act categories (similar to those of Foushee and Manos, 1998, listed in introductory section). No correlation was found for any of the communication categories. Further analyses of the original data from the Foushee et al. (1986) study of the impact of fatigue (and crew familiarity) Kanki, Lozio and Foushee, 1989, and Kanki and Foushee, 1989, did not identify correlations between distribution of types of speech acts and performance as such. However they did uncover a seemingly robust phenomenon, namely that similarity of communication patterns (crewmembers exhibit similar patterns) indicates better performing crews (see next section). In the Kanki et al. analysis of communication data, the sequences of utterances were tagged in terms as speaker and hearer and subject matter. Results showed that the better performing crews exhibited a more efficient information exchange with structuring of commands and validation of acknowledgements.

#### 2.5.3.3 Homogeneous communication patterns and performance

The Kanki et al. (1989) analysis identified *similarity* of speech patterns across different phases and types of flight (normal vs. abnormal situations) as correlated with low-performing crews. The authors interpreted this correlation as suggesting that crews exhibiting a more standard and thus more rigid and predictable communication pattern were less adaptable to very different requirements of varying types of tasks, whereas high-performance crews adapted their communication pattern to a much greater extent. In a further analysis of the same data, Kanki, Greaud and Irwin (1991) have shown that similarity of communication patterns within the crew may distinguish high-performance crews.

#### 2.5.3.4 Temporal aspect of communication and performance

In the ship bridge study of Andersen et al. (1986) the coding of the utterance-by-utterance communication involved not only the categorisation of speech into speech act categories but also the temporal aspect of the communication - i.e., whether an utterance was about past, present or future events. Results of the analysis demonstrated that the temporal aspect of communication was significantly correlated with performance outcome: ship captains who had one or several groundings or near-misses exhibited a form of communication which was less future oriented than captains who had no groundings or near misses. The authors interpreted this finding as indicating that subjects' whose technical and social resources were abundant would have a better chance of handling the current problem *and* of planning ahead, whereas resource-poor subjects often had to spend all their resources and attention on the current problem and had no spare capacity for planning ahead.

#### 2.5.3.5 Cockpit instrumentation, communication and performance

A number of studies have been made comparing communication across conventional and glass cockpit aircraft. As reported in Wiener (1993) an extensive study by Costley et al. (1989) compared three types of two-pilot aircraft: 737-200 (conventional aircraft), 737-300 and the 757 (both EFIS). The data collected allowed a comparison of rate of communication for different phases of flight (climb, cruise, and descent) and analysis showed that the 737-300 pilots were *more* communicative than those in the 737-200 whereas the 757 pilots were *less* talkative. It is therefore not surprising that commentators observe that results are hard to interpret (Wiener 1993). In a similar vein, Wiener reported on an extensive simulator study comparing DC-9 and MD-88 crews, in which effects of crew co-ordination and communication were coded. In general, the performance differences between the two aircraft types were small and on no measure did the MD-88 crews outperform the DC-9 crews, but the latter rated their own workload as lower than that of the MD-88 pilots. Analysis of the coding of communication failed to reveal differences as well.

# 3 Attention modes

*"Discrepancy between what the pilots see and they say"*

Two important issues of above statement should be considered in more detail to clarify the basis for the discussion Having in mind the central role of EPOG (Eye Point of Gaze) in understanding the individual situational awareness (SA) we need to acknowledge that we are in some way preconditioned with the assumption that what the pilot look at equals what the pilot sees. This is true only to some extent, may be during 20% of pilot working time when flying from Frankfurt to London. Most of the time the placement of gaze is not directed by the conscious attention, but simply the eyes are starring at some locations. Starring means that the eyes perform saccadic scanning without any purpose, from the point of view of acquiring the visual environment. It can  be said that the gaze is deposited at some locations without contributing to the SA. Passive staring imitates active looking, because usually when starring the eyes continue to saccade and the fixations durations not differ much from the active looking guided by conscious attention.

Conscious attention is the domain where the situational awareness is built-up. Conscious attention can be involved in three different modes of mental activity.

## 3.1 Exploration

EXPLORATION of visual environment, which dominates when performing novelty tasks like in contingency – emergency situation under unexpected circumstances. This mode is predominantly stimuli-responses operation and it operates as „real time” control. This mode dominates when performing for the first time the task, which later on due to training will became routine, the learned skill.

## 3.2 Monitoring

MONITORING the flight environment when performing routine task based on the previously learned skills. Under this mode the conscious attention is involved in the monitoring of the visual environment, to the much lesser extent as it is the case in the „exploration” mode. Conscious attention selects only the most appropriate routine for fulfilling the task. Only momentarily engagement of the conscious attention is necessary for checking the appropriateness of the used routine and the gaze is send to the relevant areas of interest (AOI) only for brief periods of time. It can be just few or even a single fixation on particular AOI, which is meaningful for checking the accordance of the model driven control responses, with the real behaviour of the aircraft. The internal model of the „real” is a basis for the skill-based control. In this mode unlike in the exploration, the reactions of the pilot are proactive – in the sense of being pro-actions based on predictions of the dynamic behaviour of the model.

Depending on how demanding the flight situation is, in terms of the required mental workload (MWL), the pilot can use the portion of conscious attention left unoccupied in two different ways. In highly demanding situations it is necessary to perform several tasks in parallel to cope which rapidly changing flight conditions. The pilot can simultaneously perform several tasks only because each of the executed routines involves the conscious attention only for brief instances in the alternating sequence.

The eye movements (EPOG) under such condition will be very much alike the eye movements during exploration. Different (AOI's) are attended and they seem to be somehow chaotic not related directly between each other. For that point of view, they differ from the EPOG strategy during exploration mode. During exploration predominantly the task relevant AOI's are attended. The limited variety of attended AOI's under such circumstances can be the result of canalised attention. The canalised attention restricts the request for the visual information to the strictly limited AOI's, quite often neglecting completely the AOI's which are crucial for flight safety. During the moments of tunnel vision the SA breaks apart. In less demanding situations, conscious attention is not guiding the eyes all the time in the purposeful manner.

## 3.3 Planning

PLANNING is the mode of the conscious attention when it is exclusively occupied various information already available in the system. There is no necessity for acquiring new visual input, as it was necessary in exploration and monitoring modes. Planning as a term is used in the broad sense to stress that the conscious brain is occupied exclusively with the future or with the past of the flight environment. There is no demand for acquiring and evaluating the present status of the flight. During this mode, usually the eyes continue to saccade, but with less intensity (accumulated saccadic amplitudes within 3 sec. time-window). It is also possible that the eyes nearly cease saccadic activity, but still there is no one continues fixation over 500ms duration. The saccadic amplitudes when passively staring on something can be as small as one degree of amplitude or even less and due to the limited resolution (sensitivity) of the eye movement-measuring device these cannot be noticed when evaluating the experimental data. There is a spontaneous reflex, which keeps the fixation duration an equal 22 ohms on average. We can extend the fixation duration over longer period (e.g. several seconds) by overriding the re-fixations reflex, when consciously attempting to look at something very carefully. This requires involvement of the conscious brain, and usually happens when a pilot is trying to aim the weapon on the target or in the civil environment when he is trying to do something perfectly like trying to perform the best ILS based approach (Instrumented Landing System). In such circumstances his visual system is focused on one single instrument in the cockpit, which is relevant to the given flight situation. This aspect of minimal saccadic intensity, can serve as the key feature enabling the experimenter to differentiate it from passive staring at some areas, which are not relevant to the flight situation.

Conscious attention in the planning mode can be involved also in some aspects of the pilot's personal life like family problems, social interaction problems, personal career, money. When in private life a pilot loses control over them, they become dangerous traps for the pilots' conscious attention, disrupting completely the situation awareness during the flight. The eyes when staring can land on any of the AOI's, without the intention to acquire the visual input from these particular locations. When the visual system is left without the orders from the conscious brain, we are usually looking straight ahead with the eyes centred in the eye cavity, close to the primary eye position. The head does not necessarily need to be in the neutral position. The rules governing the EPOG during the periods of „visual inattention", are completely different from the rules during the visual attention when the pilot is in the exploration or monitoring modes. In the first line the visual discomfort caused by bright sunlight coming from the cockpit window will likely force the pilot's eye to remain in the shadow.

Ambient vision, which is directly linked to the preconscious brain, will spontaneously direct the gaze to areas where something is changing continuously, because peripheral vision responds best to such change. The changing digital display can act as an attractor for the central vision but even under such circumstances, the content of the observed display (the reading of the instrument) is not reaching the conscious brain. The gaze shifts in the saccadic manner between purely visual attractors, without transmitting the image projected on the fovea to the conscious brain. It is hard to understand such a situation, because the central vision usually supplies visual information directly to the conscious brain. The visual information projected on the central retina can be acquired only when is requested by the conscious attention. The pilot can read the instruments only by allocating the attention on this instrument simultaneously witch the landing of saccade on this particular instrument. Passive staring excludes the attention from being at the same location within the visual environment where the eyes are just pointing. Conscious attention is than completely preoccupied with some of the already mentioned internal distracters (e.g. personal problems). The conscious attention can also be corrupted by untimely radio or intercom calls or by other unexpected events like trying to solve some technical problems, which are not of the highest priority for safe flight.

Under conditions of extremely high mental workload, the conscious brain is trying to put all reserves on problem solving. We can observe highly specific oculomotor behaviour during the instances of high mental workload. The eyes freeze for a while, when we try to solve the problem. When the intellectual task requires the use of memory, or imagination of future, we are usually looking up and holding the eyes in that strange position for the duration of the mental task. In this way, the conscious brain trays to avoid any new visual inputs which can distract the process of thinking. The „visual inattention" takes the eyes away up from the area where usually the new visual information is acquired. In this way even if some movements on the periphery of the retina will attract the ambient vision, these changes will not alarm and distract the conscious brain. When the human is in the „homo sapiens" mode in its purest from the eyes looks up „pointing up toward the cortex". There is some evidence also that depending on what kind of the task it is, if it involves spatial or temporal processing, the eyes move additionally toward the site opposite to the brain half, which dominates in that particular task.

The difficulty of the answer for the question raised in the title about the reasons for discrepancy between what they see and what the pilots say, is caused by our lack of knowledge of what the pilots really see. Knowing the location of the gaze is not equivalent to knowing about acquired visual information from that particular location. Looking for something results in seeing it only when the conscious attention is simultaneously allocated to the landing site of the saccade. During the flight, the conscious brain only temporary is engaged in active looking. Most of the time pilots are prone to passively stare at the surrounding visual scene. In the civil aviation environment especially during cruise at prescribed altitude when there is a little demand on mental work and flight become boring too some extent pilots easily falls into „visual inattention" accompanied by passive staring.

## 3.4 The skill, rule and knowledge model

This dimension of analysis serves to uncover the competence and cognitive resources of the actors and their subjective performance criteria. Humans have different modes of control of their interaction with the environment. During familiar circumstances, interaction is based on a real-time, multi-variable, and synchronous coordination of physical movements with a dynamic environment. Quantitative, time-space signals are continuously controlling movements. The automated patterns are activated and chained by cues perceived as signs, no choice among alternatives is required.

Skilled-based activities require no conscious attention or control. All activities are smooth and integrated and the senses are only directed against the aspects of the environment needed subconsciously to update internal maps of the state of affair in the environment. When this intuitive reaction to the context is no longer effective, a mismatch will be experienced between the state of affairs in the environment and the intuitive expectations of an actor. In this case, a skilled professional will normally perceive a small number of alternatives for action and efforts will be focused on search of information, which can resolve the ambiguity, that is, performance depends on active perception often calling for exploratory actions. The important point to consider here is that an expert will need no more information than is necessary to resolve the choice between the perceived action alternatives. If only two alternatives are perceived to be present, only one bit of information is needed as long as the actor is embedded in the context, even when it is a complex work environment (Rasmussen, 1991). The skill based level in the SRK human cognitive performance model is similar to the monitoring mode presented in Section 3.1

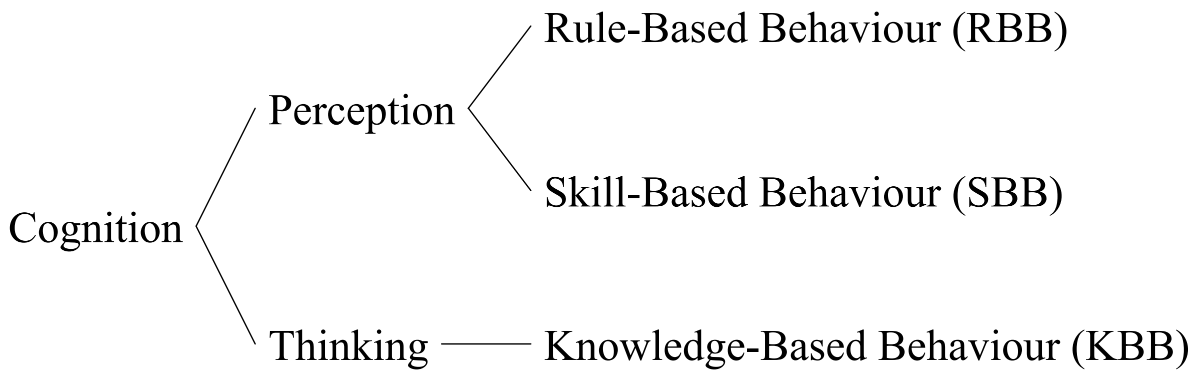



**Figure 1 The skill-, rule-, and knowledge human cognitive performance model.**

Rule-based behaviour is structured on feed-forward control through a stored rule, procedure or learned routine. Very often, the goal of an activity is not explicitly formulated but is found implicitly in the situation releasing the rule. A series of acts is often necessary to reach a certain goal. The rules may be derived empirically during previous occasions, communicated from other people's know-how as instruction or a cookbook recipe. The boundary between skill-based and rule-based behaviour in not quite distinct. It could for example depend on level of training of the operator and on the attention of the person. Since a person during a skill-based behaviour dominated situation acts without conscious attention he will not be able to describe how he controls and on what information he bases the performance. In rule-based dominated situations performance is based on know-how and an operator can report the rules used. The level of rule based behaviour is similar to the exploration mode presented in Section 3.21.
In situations where an operator is faced with an environment for which no know-how or rules are available from previous encounters, the control of performance must move to a higher conceptual level where the activities are goal controlled and knowledge based. The goal is explicitly formulated and then a useful plan is developed by selection such that different plans are considered and tested against the state of affairs in the environment. The level of knowledge-based behaviour is similar to the planning mode presented in Section 3.3
The first two behaviours are concerned with perception and action which is fast, effortless, and proceeds in parallel whereas KBB is used in analytical problem solving on a symbolic representation which is slow, laborious, and proceeds in a serial fashion. The latter one will most often be used when unfamiliar situations arise. The SRK model provides data in terms of specification of the cognitive demands of performance including knowledge and skills required, automaticity of actions and information processing requirements including perceptual and cognitive demands.

# 4 Semantic analysis and EPOG data

## 4.1 The visual unit of analysis

When the eye is stationary, and the fovea is aligned with a pattern, then one can be considered to fixate that pattern (Underwood and Everatt, 1992). Thus a fixation can be conceptually defined as: bringing fovea in alignment with a visual pattern. One single fixation must be operationally defined in time and space before being recorded, but there are no standard definitions of how to sample single fixations. For example, Just and Carpenter (1980) include in the definition of a fixation that the pattern is cognitively processed. Thus, directing the fovea towards a visual pattern without recognising the target does not constitute a fixation according to this definition. Further, if information acquisition is associated with one fixation only, the fixation definition will also depend on context variables, and consequently vary both within subject and tasks and between subjects and tasks (Yarbus, 1967). Any fixation definition must take into consideration that the analysed fixation pattern depends on the definition of one single fixation, i.e. the smallest cluster of sampling points and the sampling duration. Comparing fixation definitions that vary in milliseconds can give qualitatively different fixation patterns, or scan paths. A scan path is the pattern of lines that emerges between the defined single fixation points. Such a pattern can show sequence of fixations and how fixations are distributed in the visual field.

Eye movement tracking is not the only way of observing how subjects attend to visual stimuli. Information acquisition can also be monitored by presenting subjects with decision tasks in which they must explicitly search for information about available alternatives (Payne et al., 1978, Biggs et al., 1993). This is the same technique Kirwan and Ainsworth refer to as withheld information (1992). The problem in a full-scale simulation is that an operator can see the same information on several displays and in various representations. Operators must be able to access whatever information they want, to maintain sufficient ecological validity. Thus, withheld information may be ideal for more limited experimental studies, but less feasible for studying real tasks in complex environments. Of course, sometimes the task itself resembles withheld information, for example when one operator must ask his college about an instrument reading, simply because he does not have this information. Still, when information is normally present, the withheld information technique will not allow natural interaction with the interface.

The head-mounted lightweight infra red (IR) eye tracker seems to be a useful and ecologically valid method to measure the information acquisition of operators. Russo and Rosen (1975) argue that EPOG tracking is better than other techniques for tracing information acquisition. They compare monitoring information search with verbal protocols and suggests that EPOG tracking is more objective because it is more difficult to misinterpret. However, information search methods focus exclusively on the subjects' use of external information. It is mainly when strategies call for external information, that observable behaviour is produced.

Both eye fixations and verbal communication are process-oriented data that occur naturally. The use of joint methodologies is strongly advocated, because visual and verbal data are incomplete alone, but complementary together (Russo, 1978).

When EPOG tracking is to be used, the tasks must be arranged around gathering information (Russo, 1978). The movements themselves are very fast: typically, a 5-6 degree movement requires 30 milliseconds. This means during a typical task (consumer research), only 5 percent of the time is devoted to moving the eyes between targets, while useful fixation time accounts for the other 95 percents. To analyse eye fixations as an indication of cognitive process, one needs a model of how EPOG tracking reflects such processes. To go fishing in explorative studies is especially risky when it comes to eye movements, because fixation points are not likely to mean anything in themselves (Russo, 1978). It is essential to aggregate behavioural units, i.e. observation of single fixation points, indicating appropriate cognitive units (Russo, 1978). It is difficult to interpret information from (duration of) single fixations, because single fixations may not represent a complete cognitive unit, e.g. a morpheme referring to a graphical process parameter in the interface. Russo (1978) interpreted three or more fixations as single cognitive units.

## 4.2 Dwells on Areas of Interest (AOI)

There is no standard way of defining one single fixation operationally. One needs to define fixations in time and space, to some extent dependent on the resolution of the equipment. For example, a fixation can be defined as the three middle sample points on a surface during 250 milliseconds, or as e.g. (Optican, 1985) a rest of the eye in a location for about 150-200 milliseconds, or longer. The duration may be referred to as dwell time. The dwell time may be as low as 100 milliseconds when fixating familiar stimuli (Carl and Gellman, 1987). However, not all fixations are around 250 milliseconds. The duration of fixations upon words vary from word to word (Just and Carpenter, 1980): they used clusters of fixations on words as the smallest unit of analysis: instead of defining and analysing single fixations, they measured what they refer to as "gazes", i.e. several single fixations on a word.

Also, the factors influencing dwell time for each fixation are less straightforward. There is not always a simple relationship between the length of fixations and the amount of information obtained. Harris and

Christhilf (1980) found that pilots fixate longer on critical instruments from which information had to be extracted, rather than those requiring a mere check. Kundel and Nodine (1978) distinguished between survey dwells and longer examination dwells. Dwell time is sometimes influenced by the difficulty of information extraction. Moray and Rotenberg (1989) found that instruments were fixated more frequently after a plant failure, but that dwell times were unchanged. Russo and Rosen (1975) found few differences in dwell time between conditions in which different diagnostic strategies were used. The duration of a fixation is generally depending on context and subjects. Paulin Hansen reports fixations on fast moving objects in computer games to lie between 150 and 175 milliseconds.

Much literature deals with lower limits of fixation duration around 200 milliseconds. The visual unit of analysis should therefore be defined on a lower resolution than single fixations, with the lowest limit of duration around 200-250 milliseconds. Instead of "one single fixation", this unit will be referred to as "one dwell". The main interest in a cognitive analysis of EPOG, is the operator's use of meaningful information. This suggests seen/unseen, i.e. frequency of dwells, to be an important variable. Measures of dwell time seem to be less useful. Dwells, being several fixations within an area of interest, have a greater probability of indicating attention than have fixations. When the scenario is intensive, there should be an even greater probability of dwells being representative of operator attention. Also the areas of interest often carry detailed information (e.g. numbers) that must be accessed through the fovea (approximately 2 degrees of sharp vision). This suggests that the problem of selective attention (Broadbent, 1958 in Best, 1992) may be more theoretical than practical in such instances.

The a priori defined AOI (AOIa) has an inherent advantage in that it carries needed scenario-specific information. Furthermore, the probability of directing the line-of-gaze towards the a priori defined AOI without the intention of picking up information is low. It is difficult, and speculative, to infer how the operator use a priori AOI information based on EPOG tracking alone. For this we need to see the accessed information from the operator's perspective, namely via the verbal context. If this context is too weak to guide the choice of a match with a dwell, we can not have valid measures of voluntary visual information gathering (meaning intentional use of the information in the problem solving). However, we can still have perceptual information about operator information accessing from AOIa dwells.
By definition a "dwell" requires an area, i.e. an interface parameter, upon which the line-of-gaze is directed. Since a dwell necessarily consists of several single fixations, the target is an area rather than the exact location of a EPOG marker. To extract information from a visual target representing a process parameter, the fovea (the line of gaze) must be moved around, thus moving the line of gaze within such a target. This area will be an "area of interest". Regardless of other sampling criteria, i.e. what guides the EPOG analysis, the basic unit will be a dwell on an area of interest. (Terminology such as line of gaze, point of gaze, or fixation point could have been used, but does not emphasise the voluntary acquisition of interface information.) For a stationary eye to be categorised as a dwell on AOI, the following criteria should be met: The fovea, i.e. a line of gaze, must be directed towards an area of interest. Dwells are conceptually defined here as several single fixations within an area of interest. The operational definition is that the EPOG must remain for a minimum of 200 milliseconds inside the border of an AOI (or be within 0.5 degrees visual angle from the AOI centre, to allow for small calibration inaccuracies). As soon as the EPOG leaves the border of an area of interest, the dwell is terminated.

## 4.3 Probe events driven sampling

Probes are designed to initiate behaviour that spread throughout the team. Thus, probes can be used to study distributed teams, such as ATC. Probes make it possible to maintain a sufficient level of experimental control and at the same time have a high degree of realism, because confederates in the simulated role play / situation control the probe. Probes can be placed at rather exact point in time, e.g. directly after the DV. The benefit is that the confederates adjust the whole situation so that the probe fits in now. Probes carry with them a baseline; it will initiate a predefined normative set of behaviours that can be observed. This baseline can be developed into an operational definition of the concept of interest

The benefit of process oriented measures is that time windows can be defined anywhere, in retrospect. However, for reducing analysis time, event driven sampling is often recommended. Probes identify start and stop for a time window, thus event driven sampling will be similar for all observations. In this time window, other events than the probe itself may be studied, and they will be studied at the time of relevant changes of a situation, although they may not themselves be directly related to the probe event

The disadvantage is that events happening outside the probe event or the probe time window will not be analysed. Thus, surprising behaviour, i.e. data driven research, and possible interaction effects between probe variables and these unknown events will not be studied. However, as long as the measures are process oriented, time windows can be moved around, and at the end of the day, it may include the whole sample, i.e. the whole scenario. Thus, this is a question of analysis resources. Generally, one could say that the better prepared and well researched a new study is before it starts, the more deductive design.

You can not have a high degree of experimental control and a high degree of realism, because experimental control means reducing phenomenon, whereas realism means holism and possible (unknown) interaction effects. Analysis and synthesis at the same time is a contradiction in terms. The traditional view is therefore to first establish a causal relationship, i.e. high level of experimental control. Later, the findings are tried out in higher degrees of realism. The more real, the better the generalisations.

A modified view is that a feasible level of ecological validity is a necessary but not sufficient premise for high levels of external validity, i.e. the research goal of generalising findings. This is because real life phenomenon, often dealt with in applied research, cannot be reduced to simple laboratory settings. One need a certain level of complexity to replicate real life phenomenon in a laboratory.

## 4.4 Visualising SSA Based on Lag Sequential Analysis

Sequential analysis is a collection of techniques developed for the study of the temporal structure of sequences of events. A commonly used method is lag sequential analysis, which allows you to calculate frequencies of transitions between dwells on AOI’s within a certain lag in a time series. Lag sequential analysis allows you to answer questions like: "How many times is the a dwell on the primary flight display followed by a dwell on co-pilot's hands?" 4

The output of lag-sequential analysis is matrix with a number of transitions between behaviour and another, a so-called transition matrix (see Figure 1).

A lag-sequential analysis is relevant to the analysis of SSA because it shows how many times an act of person A is follow by a certain behaviour of person B (for details see Andersen and Pedersen, 2001). Notice that Subject B’s “response” to Subject’s A behaviour can be:

- random, i.e. it just happened to fit into person B work to behave as such, independently of what person A was doing. There is no SSA.
- practical, i.e. the team have implicitly learned that when person A behave as such, then Person B will do this in order to move on with the task or solve the problem.
- predetermined, i.e. the team must follow a specific procedure.

Before a lag sequential analysis can be made, the observation, often recorded on a video tape must be scored, i.e. it must be determined which behaviour are of interest and when do they occur during the experiment. The Observer software package from Noldus is one tool for coding behaviour.

---

4 Lag sequential analysis assumes that the transition probabilities remain constant across time: the assumption of stationarity. However, this assumption is sometimes violated, especially in long sequences of behaviour or sequences describing interaction between two or more subjects. We can inspect the degree of stationarity of EPOG data by doing an iterative lag sequential analysis for multiple time-based intervals.

In the following, we will refer to a state, which in general is the specific behaviour of interest performed by a team member. Using terminology from the Observer software, a state can consist of combinations of Actor, Behaviour, Modifier1 and Modifier2. Examples of states are: SubjectA_looks_at_Monitor or SubjectB_wawes_hand

A lag sequential analysis is the counting of transitions from one state to another. For example how many times does SubjectA_looks_at_Monitor after SubjectB_wawes_hand. (Here we only look at transitions from one state to another, i.e. using the above example SubjectB_wawes_hand must be the state immediately after SubjectA_looks_at_Monitor  to be counted as a transition. We are not dealing with case like: so many times does SubjectA_looks_at_Monitor within a period of x seconds after SubjectB_wawes_hand. Notice however that this is just two different way to perform a lag sequential analysis the proposed visualisation can be applied to either type of lag sequential analysis as the basic structure (the transition matrix) are the same for both lag sequential analysis types )

The outcome of a lag sequential analysis is a transition matrix as the one shown in Figure 1.

| | | | | | |
|---|---|---|---|---|---|
| SubjectA_behaviour1 | 0 | 3 | 11 | 12 | 6 |
| SubjectB_behaviour1 | 4 | 0 | 4 | 11 | 7 |
| SubjectA_behaviour2 | 7 | 9 | 0 | 1 | 14 |
| SubjectB_behaviour2 | 1 | 12 | 17 | 0 | 2 |
| SubjectA_behaviour3 | 5 | 5 | 7 | 8 | 0 |
| | SubjectA_behaviour1 | SubjectB_behaviour1 | SubjectA_behaviour2 | SubjectB_behaviour2 | SubjectA_behaviour3 |

**Table 4. An example of a transition matrix, which is the result of a lag sequential analysis.**

Usually more than 5 states, as shown in ***Table 4*** will exist making the transition matrix larger and more difficult to overview and interpreted.

### 4.4.1 The SSA Visualisation Method – Step 1 Re-organising the Transition Matrix

The first step to translate the transition matrix into easily interpretable information with regard to SSA is to re-organise it, so all states related to SubjectA are grouped together and similarly with states related to SubjectB. Hence the transition matrix in ***Table 4*** is re-organised as shown in **Table 5**.

| | | | | | |
|---|---|---|---|---|---|
| Subject B_behaviour1 | 4 | 4 | 7 | 0 | 11 |
| Subject B_behaviour2 | 1 | 17 | 2 | 12 | 0 |
| Subject A_behaviour1 | 0 | 11 | 6 | 3 | 12 |
| Subject A_behaviour2 | 7 | 0 | 14 | 9 | 1 |
| Subject A_behaviour3 | 5 | 7 | 0 | 5 | 8 |
| | SubjectA_behaviour1 | SubjectA_behaviour2 | SubjectA_behaviour3 | SubjectB_behaviour1 | SubjectB_behaviour2 |

Quadrants

| | |
|---|---|
| 1 | 2 |
| 3 | 4 |

**Table 5. Re-organised transition matrix for Table 4. The re-organisation makes its possible to regard SSA as indicated with the grey areas.**

Due to the re-organising the matrix can now be divided into 4 quadrants.  Each of the quadrants gives the following information:

1. How many times does SubjectA “response” to SubjectB’s behaviour. SSA initiated by Subject B.

2. How many times does Subject B perform one behaviour after another of SubjectB’s behaviour? SubjectB’s individual shifts between behaviours.

3. How many times does Subject A perform one behaviour after another of SubjectA’s behaviour? SubjectA’s individual shifts between behaviours

4. How many times does SubjectB “response” to SubjectA’s behaviour. SSA initiated by Subject A.

Based on the assumption that an indication for SSA is the number of times a behaviour of one team member initiates a behaviour of another team member, SSA can be assessed by the numbers in quadrant 1 and 4 (c.f. **Table 5**).

### 4.4.2 The SSA Visualisation Method – Step 2 Visualising

The second step of the method is make it easy to interpreted the numbers shown in quadrant 1 and 4. One solution is to calculate the sum of transitions in quadrant 1 and 4 as an immediate measure for SSA. However, the disadvantage of this approach is that the details are left out.

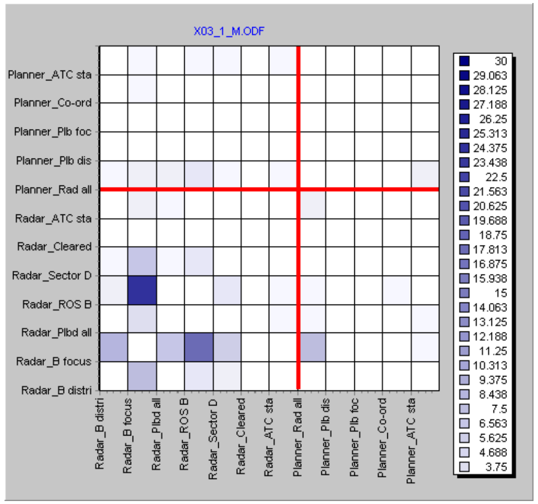
In order to maintain the details and still make it easy to interpret the many numbers, the numeric values are visualised graphically. Another benefit of this approach is that graphical patterns might appear which can be compared between teams or between different situations for the same team.

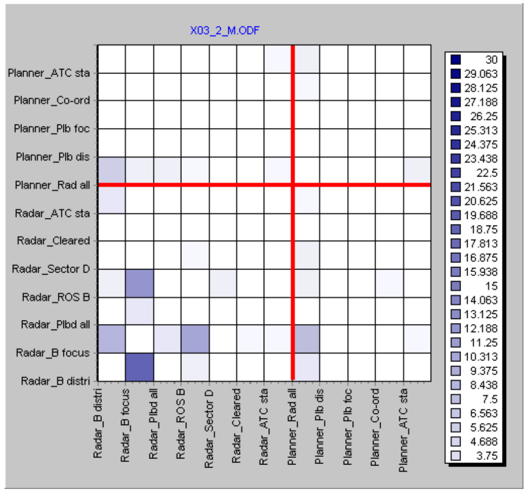
Using continuous colour coding of the numeric numbers in the transition matrix and placing the colour in a graph correspond to the layout of the transition matrix comprises the visualisation and makes pattern recognition between or within teams possible.

### 4.4.3 Applying the TSA Visualisation Method to the Air Traffic Domain

From the air traffic domain, observations on air traffic controllers have been made (see Section 6.5 for more details). A team consist of a person monitoring the radar (here named Radar) and another person working at the planner board (here named Planner). For the set of xx behaviours of interest for the Radar and yy behaviours of interest for the Planner a subset have been chosen as important for TSA. The TSA visualisation method described above was applied and the resulting charts for two teams are presented in Figure 3 and Figure 4.

Within one observation, 3 probes are placed and the team’s TSA is assessed at each probe. Probe 1 and 2 (i.e. chart 1 and 2 from the left in Figure 3 and 4) are for normal working condition and “something” just had happened in Probe 3.



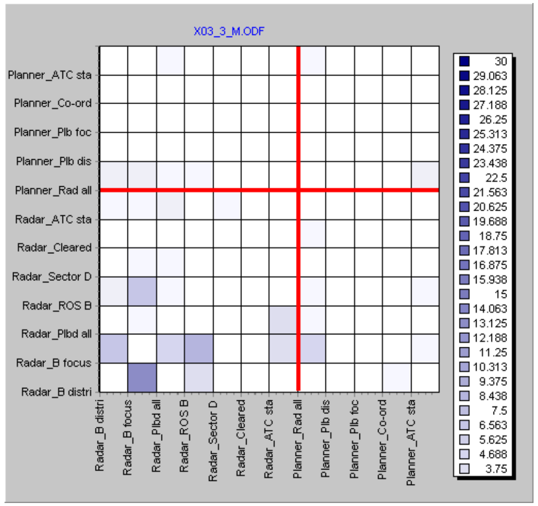
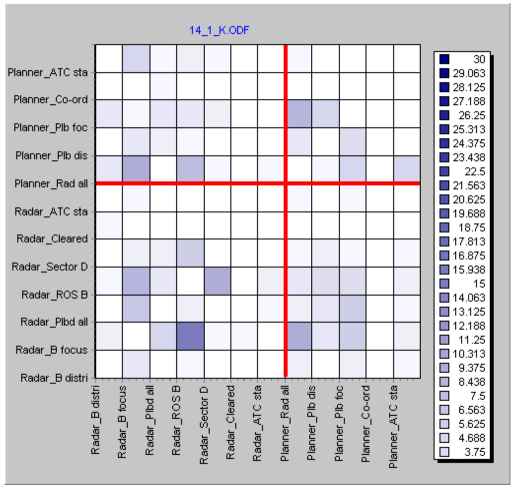
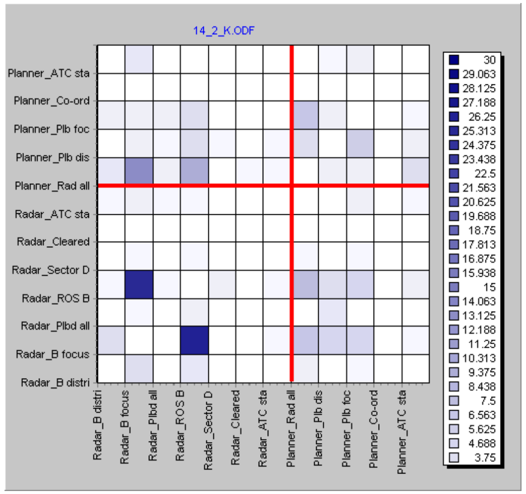
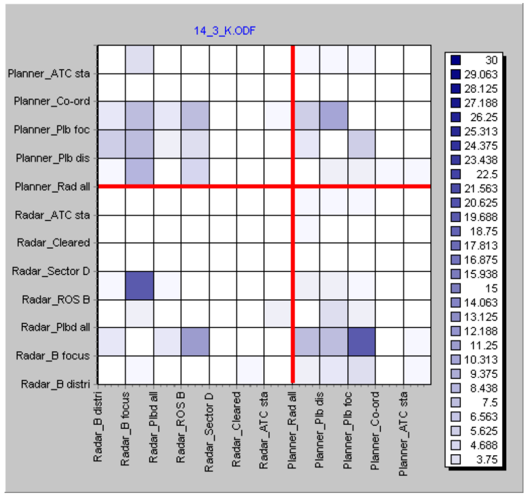




**Table 6 TSA result from team3 in the Air Traffic Control Study**




TSA result from team

**Table 7 TSA result from team14 in the Air Traffic Control Study**

Comparing quadrant 1 and 4 from ***Table 6*** with the same quadrants in ***Table 7*** more colours are observed in ***Table 7*** indicating higher level of TSA (higher number in the transition matrix). Moreover, in Figure 4, more colours appear in the chart for probe 3 (to the right) compared to the two previous probes. This indicates that when the team was in a difficult situation (probe 3) the team members paid more attention to each other.

Correlating the TSA results based on the lag sequential analysis with two the performance measurers: response time and effectiveness, the result in table 1 indicates the team14 performed better than team3.

| | response time(1-5) | effectiveness (1-5) |
|---|---|---|
| Team 3 | 3 | 3 |
| Team 14 | 5 | 4 |

**Table 8 Selected performance measures from Air Traffic Control Study. The scale range from 1 to 5 with 5 as the best.**

Hence, in this case there appears to be a correlation between TSA asses through lag sequential analysis and team performance. However, many research questions remain:

1. Is lag sequential analysis the proper way to assess TSA?
2. How do you select the appropriate behaviours relevant for TSA to perform the lag sequential analysis on?
3. Which performance criteria should be used to assess TSA?

Though this preliminary study have given some ideas for how the results of lag sequential analysis can be re-organised and visualised making it easier to interpreted the data with regards to TSA.

# 5 Combining visual and verbal data

Tracing the information accessing of subjects can not alone indicate how subjects make use of the detected information, i.e. EPOG tracking can not represent so-called higher cognitive processes without an analysis context. For this context, verbal data are needed (Biggs et al., 1993). Verbal data is unique because cognitive meaning lies inherent in the data. Thus, when studying intentions, verbal data is potentially a more direct data source than other observations of behaviour, both intentional and psycho physiological/autonomous behaviour. The same is true for estimates, i.e. subjective estimates are already processed so that they represent a statement, an inference, about the subject/the task. The main problem is to judge when verbal data and estimates are true or false (or irrelevant), and to decide what baseline should be used.

The methods do not easily allow for insight into a decision-maker's use of the information. Information acquisition methods alone provide no guaranty that the indicated information is actually processed (Payne et al., 1978). The verbal protocol may show that only limited parts of apparently accessed interface information is used in diagnosis. An optimal combination might be EPOG tracking measure of information access and VP measures of current thinking.

The ability of verbal protocols to detect meaningful thought processes represents one of the greatest advantages of protocol over eye-movement and explicit information acquisition procedures (Payne et al., 1978). The meaningful perspective of the operator can be inferred using VP. When this is the purpose, VP should be regarded as a more valid measure of the problem solving of the operator, than information acquisition procedures used alone. Finally, the information acquisition studies usually require the decision task to be more structured. This may present problems as decision researchers attempt to do more realistic research (Russo, 1978). One need to employ more than one process tracing technique and these methods should not contradict each other. Convergence is found in such multiple methods (Payne et al., 1978). Newell and Simon (1972) find high convergence in the interpretation about problem solving based on eye-movement data and on verbal protocols.

Concurrent "thinking aloud" verbal protocols could be used together with interruptive verbal protocols, i.e. interruption of a task to ask questions, to trace diagnostic problem solving strategies (Kaarstad et al, 1994, Kaarstad et al., 1995). Subjects typically stop to verbalise or verbalise in an unclear manner during intervals where the workload is judged high. At the same time, these high workload intervals are the ones that are assumed to be of interest with respect to committing various types of cognitive errors. A suggested solution to the problem is filling in the incomplete gaps in the concurrent verbal protocol by using the interruptive verbal protocol. In the interruptive verbal protocol (VPi) subjects are simply asked about the missing information (e.g. "where did you get this information"). The operators typically answer "from the interface" when asked where they accessed information. The validity of retrospective techniques, like VPi, is seriously questioned in the literature (Nisbett and Wilson, 1977). Such a merging could also be vulnerable to errors of commission (i.e. completing VPc with fabricated data from VPi). Kaarstad et al. (1994) concludes that the incompleteness problem must be better understood in order to be dealt with. At the same time it is suggested to look into eye movements as an aid to reduce the seriousness of the incompleteness problem in VPc.

Verbal data can be used in explorative studies, i.e. hypothesis generation, building and testing models of cognition, or to supplement other data as part of the analysis (Payne et al., 1978). Payne et al. (1978) mentions an example where monitoring of information acquisition behaviour and collection of verbal protocols where combined: the decision task is set up so that the subject must view or select information in a way that can be easily monitored. However, EPOG tracking may be a better method within process control, because it is: concurrent with problem solving, potentially more accurate than other techniques, ecologically valid an can easily be registered together with VP. Others have used EPOG tracking to study cognition (Winikoff, 1967, Just and Carpenter, 1980, Russo and Rosen, 1975). However, little research has been done on how incomplete concurrent protocols can be clarified using EPOG tracking.

In tasks where operators alternate between intentional visual information acquisition and verbal communication, one would expect a strong relationship between input and output of the information processing system. The point of gaze may represent an input to the information processing of an operator. This input could be the result of a top down decision to search for visually presented information, or it could be a bottom up process where the operator decides that a signal stands out from other signals in the visual field of interest.

# 6 Experiments

## 6.1 Simulated Flight Task

Situational Awareness (SA) is recognised as an important factor in skilled performance and effective decision-making in various dynamic environments, including the aviation flight deck, and air traffic control. Both experimental and operational data point to the importance of SA in decision making and effective and accurate assessment of the operational environment. The ability of the pilot to maintain situational awareness (that is his ability to monitor and understand the state of aircraft, its systems and its environment) is crucial to mission success and, last but not least, survivability. That means that maximising the pilot's SA is essential for ensuring optimal pilot performance (Endsley, 1988).

Although the importance of SA seems obvious, there are still many opinions about how to define SA. The most widely known and most often used definition of SA is that of Endsley, who defined SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of

their meaning and the projection of their status in the near future" (Endsley, 1988, p.789). According to this model, SA consist of three components (levels):
- 'Perception of elements in current situation' (a pilot must perceive the radio messages, changes on the displays indicators of his own plane, actions and characteristics of other aircraft etc.).
- 'Comprehension of current situation' (the perceived information should be synthesised and interpreted by the pilot)
- 'Projection of future status' (this can be seen as the highest level of SA, which means that a pilot should make the predictions of future events on the basis of his interpretation of the information.
According to Endsley (2000), the relationship between SA and performance implies that a high degree of SA will considerably increase, but not always guarantee, the probability of the right decisions and of good performance. On the other hand, lower SA results in both errors and wrong decisions. According to Jones & Endsley (1996) we can speak of three major categories of errors that can occur in aviation which are based on the three levels of Endsley's definition of SA:

- Level 1 (failure to correctly perceive the information);
- Level 2 (failure to comprehend the situation); or
- Level 3 (failure to project the situation into the future).

The Aviation Safety Reporting System database used by Jones & Endsley showed that the largest percentage of errors (76.3%) was Level 1 SA errors (incorrectly perceiving information). From aircraft accident investigations it is known that it is often difficult to discover why exactly an operator (e.g. pilot was not aware of some critical factors because there are too many factors that may affect awareness level (Rodgers, Mogford and Strauch, 2000). The workload and communication between crewmembers are two important factors, which can play a significant role in achieving maximum performance and a high level of SA in a cockpit.

Many investigators consider the relation between SA and workload critical (Taylor, 1990; Adams, Tenney, and Pew, 1995; Brookings, Wilson, and Swain, 1996; Endsley, 2000; Wickens, 2001). Generally, when (mental) workload increases too much, perception will degrade. This can result in the pilot not perceiving changes in the environment, reacting too late to such changes, or reacting when it is not necessary.

Besides the degree of workload, communication and co-operation between team members also play an important role in performance. Performance in many complex systems depends on the co-ordinated activities of a team of individuals (Cannon-Bowers and Salas, 1990, as cited in Salas, Prince, Baker, and Shrestha, 1995). Teamwork consists of behaviours that are related to team member interactions and are necessary to establish co-ordination among the individual team members to achieve team goals (Salas et al., 1995). As noted earlier, situational awareness plays an important role in individual performance. The concept of SA can also be applied to the performance of a team. Endsley (1995) defined team SA as "the degree to which every team member possesses the situational awareness required for his or her responsibilities" (p. 39). To share awareness with respect to a situation means that all crewmembers perceive and comprehend the situation in the same way, and they know that they share this awareness of the situation. SA of a crew could suffer if, for instance, one or more crewmembers failed to detect or rectify errors because of a lack of awareness, and nobody else gave a warning. Good shared SA, just like good teamwork in general, therefore involves such activities as co-ordination and information sharing. Good co-ordination often means good communication. Verbal and non-verbal communication with others, even in situations with restricted visual cues (like the work of air traffic controllers) has been found to be an important source of SA information (Endsley, 2000). Schwartz (1990, as cited in Salas et al., 1995) pointed out that the level of crew situational awareness was related largely to the level and quality of communication observed in the crew. Incomplete communication between crewmembers was seen as an indicator of decreased situational awareness. Besides this it can be assumed that when workload is too high it can be a threat for the quality of the co-operation and communication between team members, which can result in reduced sSA.

As mentioned before, in the study of Jones & Endsley (1996), the incorrect perception of information is probably the most common error in aviation. To investigate visual perception in general and specifically that of pilots, there are many different ocular indicators currently available to researchers. One of these is the Eye Point-Of Gaze (EPOG). This is a physiological measure which allows real-time registration of the pilot's scanning behaviour. The data obtained from EPOG include among other things the number and duration of fixations, dwell time (the time spent looking at one bounded area), dwell percentages (fixation time on a particular area as a percentage of total scanning time), the transition rate (the frequency of the glance changing from one bounded area on the surface into another), the blink rate and the pupil diameter. In the research of the EU Fourth Framework project VINTHEC (Visual Interaction and Human Effectiveness in the Cockpit) the intention was to try to develop a method on the basis of EPOG in order to improve the understanding and judge the SA of pilots (VINTHEC WP1). The results demonstrated that the role of EPOG in this kind of research is very valuable. This method can be especially beneficial in helping to understand the first level of the SA concept ('Perception of elements in the current situation').
Aspects of scanning behaviour relate directly to mental workload (Harris, Glover, and Spady, 1986). Looking at the changes in scanning behaviour, which are important in this study, it is possible to see the influence of mental workload on the first level of SA. Workload increase can result in a decrease of the number of fixations and transitional activity, together with an increase in dwell time (VINTHEC WP6). It has also been found that an increase in mental workload is attended by a change (decline) in blink rate (Holland & Tarlow, 1972; Tanaka & Yamaoka, 1993; Veltman & Gaillard, 1998), and a dilation of the pupil diameter (Matthews, Middleton, Gilmartin, Bullimore, 1991; Hilburn, 1996).

### 6.1.1 The present study

Thus far the VINTHEC II project has focussed on individual SA (i.e., of one pilot). The general aim of the present study was to acquire information concerning the possibility of measuring crews' (i.e., multiple operator) shared Situational Awareness (sSA) in an aeroplane cockpit, and to demonstrate the utility of eye tracking measurement in this regard.
To influence sSA, two independent variables were experimentally manipulated: task load and team interactivity. This experiment started from the assumption that the quality of co-ordination between crewmembers could serve as an indirect indication of shared SA. The influence of workload on the quality of this co-ordination was also traced. During the experiment, pairs of test subjects had to work together as a team ("captain" and "co-pilot") performing tasks resembling actual flight tasks, using the Multiple Attribute Task battery. The Multiple Attribute Task battery (MAT) is multi-task flight simulation software package that runs on a PC (Comstock & Arnegard, 1992). Participants were responsible for performing different subtasks, but their goal as a pair was to obtain the best results (in other words to work fast and to try to make as few as possible errors). Both task performance and visual scanning behaviour of the participants was registered.

It was predicted that as task difficulty increased, participants would spend less time co-ordinating their actions, which would decrease sSA. It was hypothesised that visual scanning behaviour could confirm this. That is, aspects of scan pattern should differentiate between "co-operative" and "solitary" modes of team interaction, and between levels of task load. It was assumed that as an indication of rising mental workload, pupil diameter of both of the team members should also increase. It was further assumed that increased visual attention demands, which were also associated with high task load in the present study, would lead to a decrease in the blink rate. It was also hypothesised that excessive mental workload would result in poor task performance of both crew members, and that co-ordination between crew members would cause changes in the scanning behaviour of the captain: a longer dwell time, a larger transition rate and a greater number of fixations by the captain on that part of the display corresponding to the co-pilot's tasks. With respect to the performance of the team, it was expected that co-ordination would lead to better results for the co-pilot's tasks, but also to decreased (or unchanged) performance of the captain. Increases in pupil dilation as a result of high workload were expected, especially during the co-operative way of working on the tasks. Finally, it was expected that high task load would influence team interaction in such a way that transition rate and dwell time for the captain would both decrease, for that part of display corresponding to the co-pilot's tasks.

### 6.1.2 Method

#### 6.1.2.1 Subjects

Test subjects were 24 students, from various educational backgrounds (university and technical education), and ranging in age from 20 to 38 (mean = 25.1). Nineteen of these were men, and five women. Only one subject was left-handed. From these 24 test subjects, a total of 12 two-person crews was formed. Most of the subjects (22 of 24) had no prior experience with the MAT task. Participants were randomly assigned to either the Captain or Co-pilot position.

#### 6.1.2.2 Apparatus and Stimulus Material

##### 6.1.2.2.1 The equipment

EPOG was measured by means of the GazeTracker ® system. The GazeTracker consists of three subsystems: An eye tracker, a head tracker and an integration sub-system. The eye tracker is an ASL 4000SU, which is manufactured by Applied Science Laboratories. This system determines the pupil-to-cornea reflex angle using the "bright pupil" method. The corneal reflection, necessary for the bright pupil method, is obtained using an infrared LED whose beam in directed coaxial with the viewing axis of the pupil camera (the helmet mounted camera filming the pupil). An image of a bright pupil can be seen on the photographic inset in Figure 2. This bright pupil image is processed by a dedicated PC and results in a data stream to the integration sub-system. Manufacturer's specifications claim a visual tracking range of 50 (horizontal) by 40 (vertical) degrees and an update rate of 50 Hz.

The head tracking subsystem (Flock of Birds, by Ascension Technology, Inc.) relies on measurement of magnetic field disturbances to determine position and orientation of the subject's head (i.e. eye). A fixed reference transmitter, affixed to a stationary surface near the subject, emits a pulsed DC magnetic field. The receiver is mounted to the eye tracker helmet (see Figure 1). The combined head- and eye tracker system allow the test subject completely free head movements. According to the manufacturer's specifications, transmission range is 0.9 m. radially, with an accuracy of 0.3 cm. for position and 0.5 degrees for orientation, with an update rate of 100 Hz.

**Figure 2: The GazeTracker helmet assembly. On top of the helmet is a socket for the magnetic receiver. Immediately in front of the left eye is the visor, which reflects the infrared light while the person looks through the visor**

Data streams from eye- and head tracker fuse in the integration subsystem resulting in an ASCII file containing:

- Fixation number
- Surface number (up to 30 viewing planes may be defined)
- X and Y co-ordinate relative to a viewing plane (0.1 mm.)
- X, Y and Z co-ordinate of eye position in space (0.1 mm.)
- Pupil diameter (arbitrary units)
- Current time (ms.) since beginning of the day
- Fixation duration (ms.)

The sampling rate of the entire system was determined by the "slowest" sampler of the system: the pupil camera, which samples at 50 Hz. The accuracy of the entire system is claimed by the manufacturer to be 1 degree. The accuracy of the system can, to a certain level, be set according to experimental needs. Higher accuracy may deliver more information but is more time-consuming during equipment calibration and data analysis.

*A schematic representation of the equipment can be seen in*

***Figure 3***.

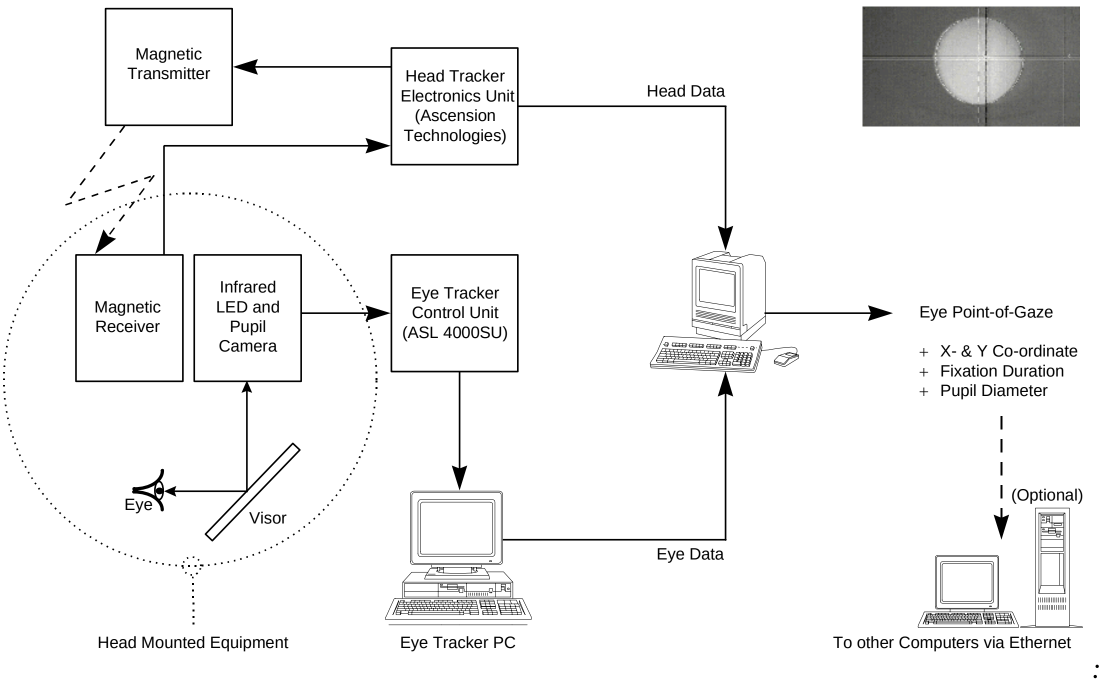



**Figure 3 Schematic representation of the GazeTracker.**

For the EPOG data collection, the space in front of the participants (including two computer monitors) was divided into 5 areas of interest. The following information was received from the EPOG equipment and recorded:
- Fixation number on areas of interest
- Fixation duration
- Fixation transitions

- Dwell time on areas of interest
- Pupil diameter
- Blink rate
- Blink duration

#### 6.1.2.2.2 Flight Simulation task

The Multiple Attribute Task (MAT) battery was presented via a simplified desktop PC and consisted of four sub-tasks, each of which taps a different aspect of cognitive-motor performance: Tracking, System monitoring, Radio Communications and Resource management. As used for this study, the MAT consisted of the following three component tasks: Tracking, System monitoring and Resource management.
Figure 3 gives an example of the display with the MAT battery used in this experiment. For a complete description of the tasks see appendix B.

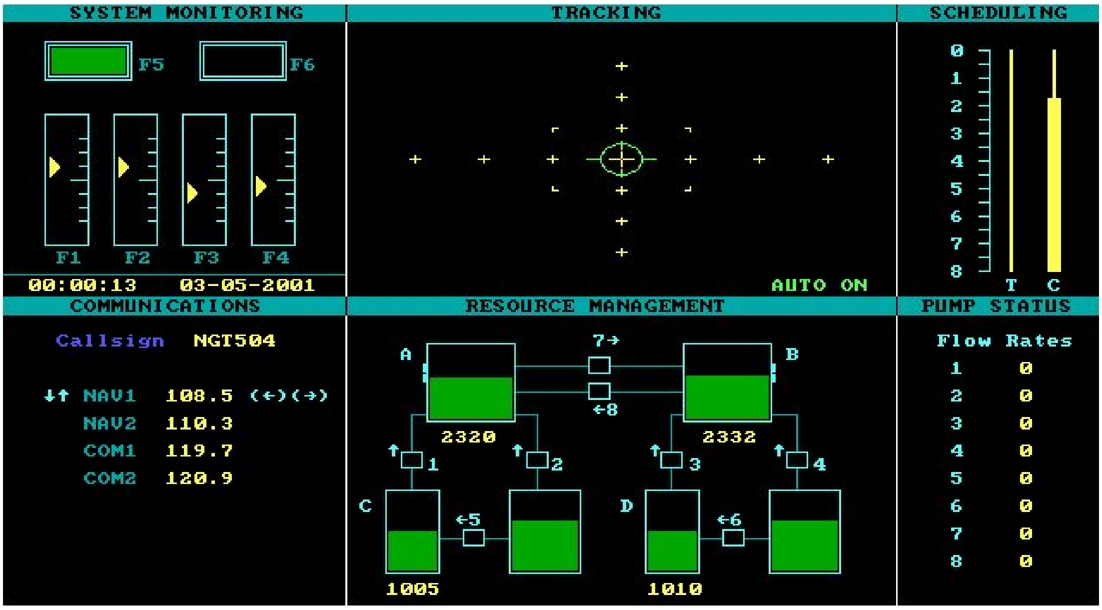



*Figure 4: The Multiple-Attribute Task (MAT) battery windows*

The MAT was run on a single PC. A video splitter permitted the simultaneous display of the MAT on two side-by-side 17-inch colour monitors. Task inputs for the co-pilot were made via a single joystick (for the tracking task), and the numeric- and function-keys of a single keyboard (for the monitoring task) and the captain used the mouse (for the resource management task).
The purpose and the descriptions of the tasks were presented to the participants by means of written instructions.

### 6.1.2.3 Design

A 2x2 within subjects repeated-measures design was used (see figure 4). The first independent variable refers to Team interactivity (Co-operative versus Solitary) which was manipulated by presenting subjects one of two alternative sets of instructions: the first set designed to stress overall team performance, and thereby encourage partners to work together, and to oversee the performance of the partner. The second set of instructions asked subjects to focus on their individual subtasks, without regard to the partner's performance.
The second independent variable refers to Task load (Low versus High) and was varied by scripted changes in task difficulty during the test sessions. Levels of task load were defined on the basis of number of system monitoring events and the sine wave amplitude of the tracking task forcing function. The High Task load

script contained 48 events per 5 minutes, whereas the Low Task load script contained 14 events per 5 minutes. As with Team interactivity, Task load was varied within subject.

Dependent variables for this study consisted of dwell time (per screen), Eye Point of Gaze, and objective task performance. This last aspect was defined as:

- Tracking RMS error;
- Monitoring response time;
- Hit rate and false alarm count (in the "System monitoring" subtask); and
- fuel management average RMS error (in the "Resource Management" subtask).

Both Interactivity (Co-operative versus Solitary) and Task load (High versus Low) were balanced in order to control for a learning effect. This resulted in four possible variations: High/Co-operative, Low/Co-operative, High/Solitary, and Low/Solitary. Those variations formed the basis of four sessions (runs) of the experiment, by which each of the teams got its own unique order of runs (see appendix A).

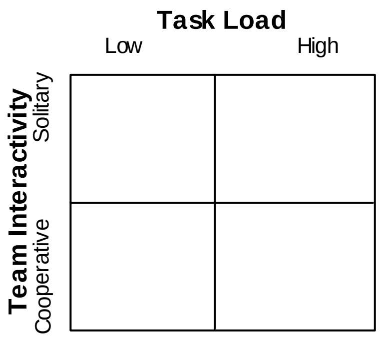



**Figure 5 Experimental design (2x2 within subjects)**

#### 6.1.2.4 Procedure

The experiment took place at the NLR (National Aerospace Laboratory) in Amsterdam. After each pair of participants arrived in the testing room, one of the team members was randomly assigned the function of 'co-pilot' (which involved performing two subtasks, "Tracking" and "Resource Management") and the other the function of 'captain' (which required performance of the "Resource Management" subtask, as well as overseeing the work of the co-pilot). Each of them took a seat behind one of the two computer screens, about 70 centimetres from the screen. Participants were seated side-by-side, approximately 1 meter from one another, while the captain and co-pilot remained on the same place.
An experimenter noted the date, time of the experiment, and the gender of the participants and the briefing questionnaire was handed out.
Before the beginning of the experiment, participants were familiarised with the experimental goals and procedures, by means of written instructions (for the complete text of the instructions see appendix B). Familiarisation was followed by 5 minutes of hands-on training.

Prior to the first session, the fit of the helmet was assessed on the subject's head. The EPOG equipment was then calibrated. This took about 10 minutes per subject.

The experiment consisted of four sessions (15 minutes per session). In two of the four sessions, the team had to work together, which meant that the 'captain,' in addition to performing his/her own task, had to monitor the work of the 'co-pilot', and if necessary assist him/her. In the other two sessions of the experiment, the 'captain' and the 'co-pilot' worked separately from one other and had to concentrate on their own tasks. Before the beginning of each session, participants were informed by the experimenter whether the coming session required them to work separately or together. Task load was also varied over the four sessions.

To determine whether any shifts occured in the EPOG calibration during the session, the calibration had to be checked, which meant that subjects were required to fixate four calibration targets at the end of each of the four sessions. Between the second and third sessions, there was a 10 minute break.
After the last session was finished, the GazeTracker helmet was removed from each participant's head. Participants were then thanked for their participation and debriefed on the project before leaving the testing room.

## 6.2 Results

### 6.2.1.1 The results of objective task performance

Objective task performance in this study was assessed in terms of the number of false alarms, hit rate and response time in the "System monitoring" task, tracking RMS error (mean deviation from centre of tracking) in the "Tracking" task; and fuel management average RMS error (absolute deviation from the goal level, 2500 units of fuel, averaged for the two target tanks) in the "Resource Management" task.

A two-way repeated measures Analysis of Variance (ANOVA) was performed on the factors Task load and Team interactivity. The results, across all teams, showed a significant main effect of Task load on the performance of the co-pilot for the "System monitoring" task (mean number of false alarms, $F(1,11) = 16.44$, $p < .05$), and "Tracking" task (tracking RMS error, $F(1,11) = 70.18$, $p < .05$). This effect on the performance of the captain in "Fuel management" task was also significant (fuel management average RMS error, $F(1,11) = 39.01$, $p < .05$). Figure 5, 6 and 7 illustrate this effect: high Task load was associated with a significantly increased false alarm rate, Tracking RMS error, and fuel management average RMS error.

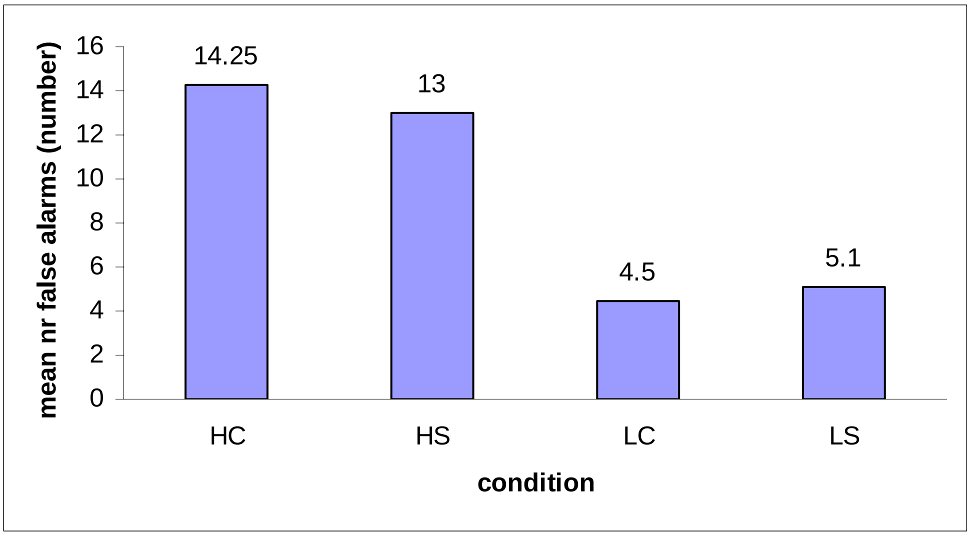



**Figure 6: The mean number of false alarms (Monitoring task) over four conditions (High/Co-operative, High/Solitary, Low/Co-operative, Low/Solitary**

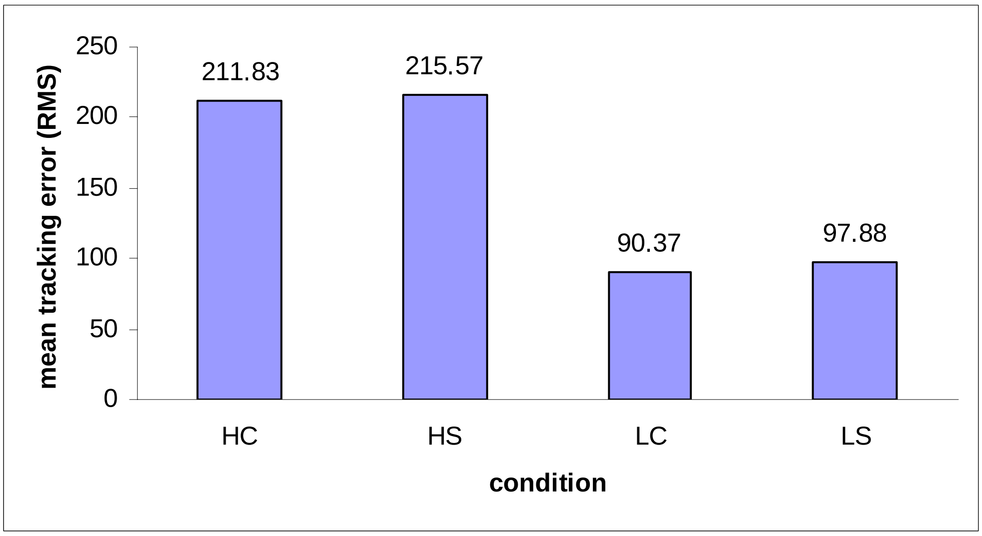



**Figure 7: The mean Tracking RMS error (Tracking task) over four conditions (High/Co-operative, High/Solitary, Low/Co-operative, Low/Solitary).**

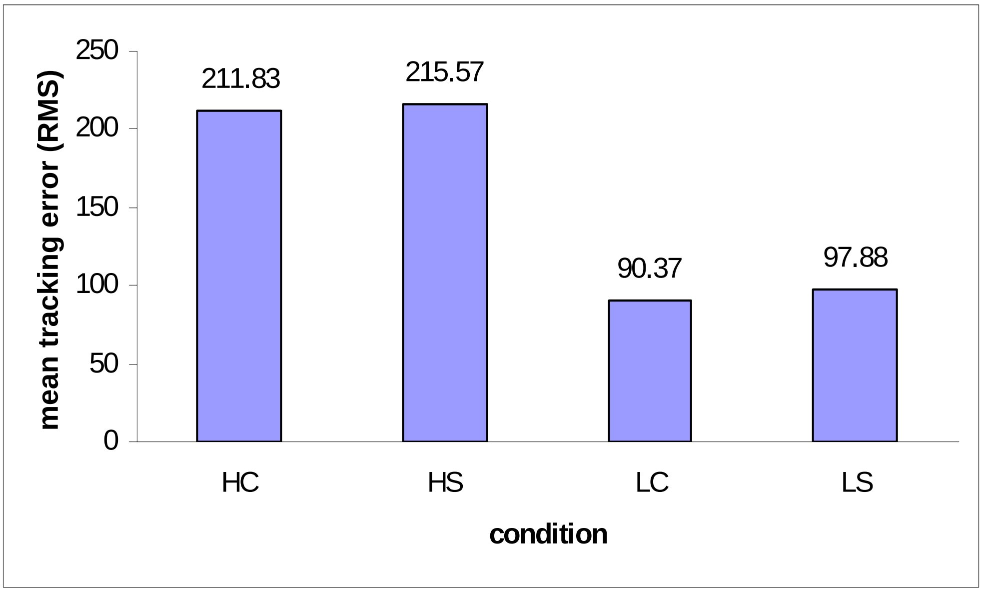



**Figure 8: The mean fuel RMS error (fuel Management task) over four conditions (High/Co-operative, High/Solitary, Low/Co-operative, Low/Solitary).**

**Table 9** gives the means and standard deviations of the task scores in the four Interactivity by Task load conditions.

**Table 9 Mean hit rate, False alarm count, Reaction time (Monitoring task), mean fuel management RMS error, and mean tracking RMS error in four conditions (High/Co-operative, High/Solitary, Low/Co-operative, Low/Solitary).**

| | Monitoring | | | | | | Fuel management | | Tracking | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Hit rate | | False alarm | | RT | | RMS error | | RMS error | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| HC | 0.97 | 0.04 | 14.3 | 13.6 | 2.04 | 0.7 | 255.2 | 100.8 | 211.8 | 81.6 |
| HS | 0.97 | 0.02 | 13.0 | 8.6 | 1.91 | 0.4 | 297.6 | 168.0 | 215.6 | 73.7 |
| LC | 0.97 | 0.04 | 4.5 | 3.7 | 2.01 | 0.4 | 87.8 | 37.7 | 90.4 | 32.4 |
| LS | 0.97 | 0.03 | 5.2 | 3.9 | 1.96 | 0.5 | 85.3 | 66.6 | 97.9 | 34.3 |

The main effect of Team Interactivity was not significant for any of the dependent measures, although a trend was found toward decreased tracking error and increased monitoring hit rate under co-operative work.

#### 6.2.1.2 Results of the EPOG measurement

The data of some of the teams was partially (five runs from three teams) or completely (four runs from one team) omitted from the analysis because less than 60% of this data could be explained. "Explained" in this context means that more than 40% of the data consisted of either eye blinks or fixating an indeterminate area. The rest of the EPOG data were analysed using a one-way ANOVA. In order to calculate mean pupil diameter z-score transformations were performed on the raw pupil diameter data, whereby Z score = (raw score - mean) / standard deviation. The rest of the EPOG data were analysed using a two-way repeated measures ANOVA. In order to calculate mean pupil diameter, z-score transformations were performed on the raw pupil diameter data, Z score = (raw score - mean) / standard deviation.

Results revealed a significant effect of the Team Interactivity on a number of parameters of the eye tracking data: dwell time, fixation duration and number of fixations of the captain on the Monitoring window, $F(1,10) = 16.72$, $p < .05$ ; $F(1,10) = 16.72$, $p<.05$ ; and $F(1,10) = 17.56$, $p<.05$, respectively. **Table 10** shows that different modes of team interaction led to corresponding differences in number of fixations and average dwell time of the captain on his own window (Resource management) and on the windows of the co-pilot (Monitoring and Tracking).

**Table 10: Comparison of the captain’s dwell time (%) and number of fixations between the co-operative and solitary modes of team interaction**

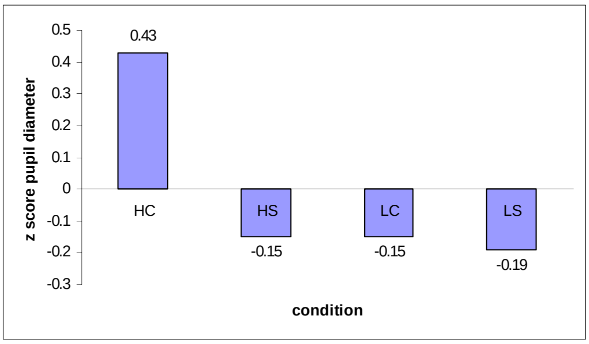
| | **Solitary** n = 20 | | **Co-operative** n = 19 | |
|---|---|---|---|---|
| | mean | SD | mean | SD |
| **Monitoring + Tracking** | | | | |
| Dwell time (%) | 9.64 | 19.6 | 19.4 | 14.7 |
| Number of fixations | 173.7 | 348.17 | 328.26 | 236.9 |
| **Fuel Management** | | | | |
| Dwell time (%) | 73.25 | 27.2 | 64.83 | 12.1 |
| Number of fixations | 1230.35 | 473.1 | 1120.11 | 270.5 |

Note: “Dwell time” refers to the percentage of time that a subject looks at a specific area;

n – number of cases.

Further there was a significant main effect of Team Interactivity on the transition rate of the captain, $F(1,10) = 19.76$, $p<.05$.

An unexpectedly small number of the scanning behaviour parameters was affected by the Task load. Under both high task load and low task load, the captain's mean dwell time on his own window (Resource Management) was significant longer than that on the co-pilot's windows. Pupil diameter of both team members showed a significant main effect of Task load, $F(1,15) = 8.09$, $p<.05$. Particularly with the co-pilot this effect was significant, $F(1,7) = 5.7$, $p<.05$. For comparison of the z-scores of pupil diameter between captain and co-pilot see figure 8.



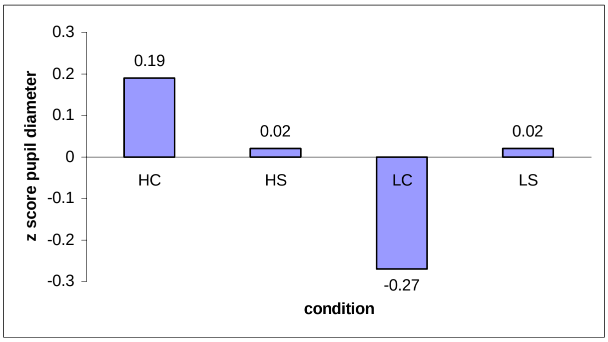




**Table 11 Comparison of the captain's (left) and co-pilot's (right) z-scores of the pupil diameter over four conditions (High/Co-operative, High/Solitary, Low/Co-operative, Low/Solitary).**

## 6.3 Conclusion

The aim of the present study was to demonstrate the utility of eye tracking in the measurement of flight crews' shared Situational Awareness. Because of the assumption that the quality of co-ordination between crew members could serve as an indirect indication of shared SA, the effect of workload on the quality of this co-operation between the team members was also traced in this study.

The obtained data consisted of two parts: the objective performance of test subjects on the tasks and the EPOG measurement data.
The result of task performance showed that the effect of the task load was significant on most aspects of this performance. This means that a high level of the task load resulted in worse performance. No beneficial effect of co-operative interaction was found on task performance of co-pilot (although a trend was found toward decreased tracking error and increased monitoring hit rate under co-operative work).
Based on the results from the EPOG measurement it could be concluded that there actually were different modes of team interactivity. When participants had to work together, there were significantly more and longer fixations, and a longer dwell time of the captain on the windows of the co-pilot compared with the situation when both team members had to work individually.
For both team members, the dilation of pupil diameter was affected by task load. The enlargement of the pupil diameter, because of high task load and the co-operative way of working, indicated the increased mental workload of team members.

In summary, it can be concluded that task load manipulations resulted in significant changes in objective task performance measures (e.g., tracking RMS error, monitoring false alarm count, and fuel management RMS error), as well as one of the eye tracking measures (pupil diameter). Pupil diameter was sensitive to task load changes for both captain and co-pilot. Further, experimental manipulation of Team Interactivity resulted in significant changes in several eye-tracking parameters (for the captain only): dwell time, fixation duration, fixation count and transition rate. The fact that co-operative team interaction had no convincing beneficial effect on the performance of the co-pilot, and an increasing of his mental workload during co-operative work, suggests that the co-pilot did not benefit from this co-operation. The reason might be that the test subjects were not real pilots and before the experiment started, they did not have any experience in communicating with each other regarding this kind of experiment.

This study demonstrated the potential benefits of EPOG measures in assessing sSA. Whereas workload effects in this study were generally observable via task performance (as well as such EPOG measures as pupil diameter and blink rate), team interactivity manipulation was more reflected in EPOG measures, and generally not observable via task performance changes. This argues for the use of EPOG measures to help infer changes in shared SA. That is, although SA appears an essential element of task performance (Endsley, 1988), performance itself is not necessarily a sensitive indicator of SA. Further studies can perhaps help explain the complex relationship between SA, workload and task performance.

## 6.4 Nuclear reactor

It is typical of complex technical real time systems that they require, for the sake of safety and efficiency, more than one operator. When two or more operators are controlling a process, the collaborating collection of operators can be referred to as a team, i.e. the concept of team situation awareness should include inter-personal aspects of awareness. So, relative to each individual operator, his or her current model of the task domain will *also* include how other team members perceive and understand the situation and how they understand *his* current knowledge. Thus, in a sense team situation awareness is a component of individual situation awareness: operators must to some extent be aware of each other's tasks and of each other's awareness of those tasks. In short, therefore, team situation awareness involves team members' *mutual knowledge* about their task domain. (See Andersen & Andersen, 2000 for details on this point).

SA may be assessed by either subjective measures (involving operators' or expert observers' qualitative evaluation of performance and behaviours) or objective ones directed at subjects' responses and task directed behaviours. In this paper - and in the study we refer to - we have concentrated efforts on objective measures- For several reasons we prefer to apply measures that do not involve an interruption of operators' task. That is, we have focused on assessment methods that involve measures that are made continuously across the evolvement of the task scenario. In addition, we have sought to define a set of ideal behaviours and allowing us to compare this norm with the observed behaviour. Thus, we hypothesise that the degree of correspondence (i.e., in terms of percentage) between a pre-defined ideal behaviour and the observed behaviour may can constitute a measure of situation awareness. As will be explained below, the team aspect will then be added to this measure by including behaviours, which involve the perception of, and interaction with fellow team members.

In this study we sought to develop integrative methods combining eye-movement tracking data with elements of, first the Cognitive Systems Engineering framework developed at Risø National Laboratory (Rasmussen et al, 1994) for analysing of operators' cognitive activities and second sociological frameworks for the analysis of everyday social non-verbal communication modalities (Andersen, 1997). The study described in this paper is a pilot study, which seeks to establish the feasibility of applying these methods of measurement and analysis, not their validity. The initial ideas for a continuous measure of team situation awareness were tried out in a small technical pilot study conducted in the nuclear reactor control room at Risø (a 12 MW research reactor). The Risø reactor control room is a naturalistic operational environment where the operators have to co-ordinate their tasks to achieve the desired level of safety and efficiency.

### 6.4.1 The experimental set-up

The data derived from the pilot study described here are currently in the phase of analysis and will be subjected to the integrative analysis described above. Since the data recordings associated with this type of are rather complex in themselves. we need to describe them briefly.

Risø invested in new laboratory equipment in 1999 enabling faster, and more flexible analyses of many types of data, visual data in particular. The use of these data recording and analyses systems is highly skill based and the set-up of components in the field is not straightforward. It was therefore decided to explore the usability of major components of the new laboratory equipment in the field with the specific consideration in mind that these types of data may be used for the analysis of team situation awareness.

**Figure 9  A nuclear control room operator wearing the eye tracking helmet.**

The two main objectives of this technical pilot study were to (1) gain hands-on experience with the eye tracking and the analysis systems under conditions where this equipment had to be operated in the field during real (non-interruptible) scenarios and (2) to acquire experience about the implementation of the suggested measures to be combined into an assessment of team situation awareness.

The selected target task in the nuclear research reactor (configuration of neutron flux by inserting and removing of neutron absorbing rods) is typically performed at intervals of 48 hours during normal operations. The exchange of rods aims at optimising the configuration of the core. The task, which requires one man at the top of the reactor to adjust the rods, and one man in the control room to monitor instruments, takes about 2-5 minutes (excluding preparations). These two team-mates have to co-ordinate their tasks closely in order to adjust the reactor in a safe way. The removal of the rod has be done very smoothly and not too fast. Failing to do so will cause the reactor to shutdown automatically. For the pilot study data were collected through:

- Combined head- and eye-movement tracking from the operator in the control room.
- Video recordings of the operator at the top of the reactor.
- Video recordings of the 3D model of the instrumentation in the control room (mixed with eye point gaze data)
- Audio recordings from both operators.
- Questionnaire.
- Debriefing with the group of operators

***Figure 10*** illustrates the overall set-up in the reactor control room with one of the operators and one of the researchers. The second operator is located on the top of the reactor, but is visible to his college through the monitor in the control room.

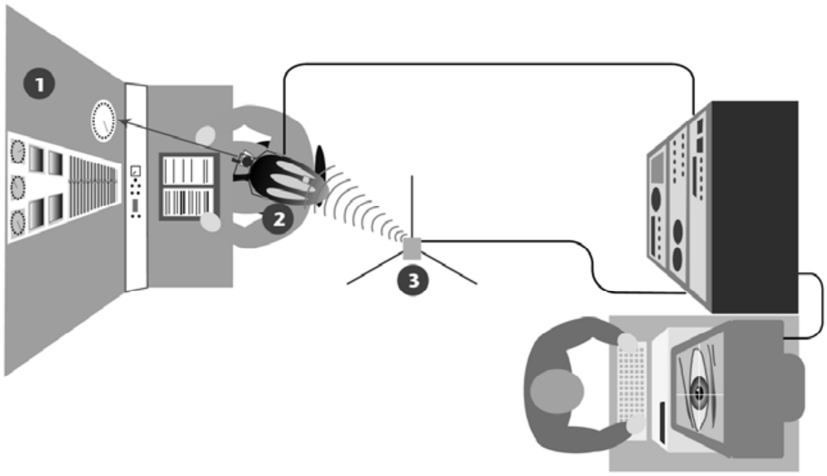


**Figure 10 The technical set-up of equipment.**

(1) The reactor instruments, monitored only by the operator in the control room, display how the neutron flux in the reactor core is changing as a consequence of the team-mate's manual removal of neutron absorbing rods. The operators communicate through the intercom.

(2) The operator wears the eye-tracking equipment. An effector placed on the top of the helmet gives the position of the head relative to the environment using a magnetic field. Together these enable continuous measurement of what instruments the operator is looking at. Both operators are wearing wireless microphones.

(3) The magnetic tracking system combined with a laser pointer tool is used for building a 3D computer-model of the control room. The magnetic transmitter is the reference point for the effector mounted on the helmet, enabling integrated eye- and head tracking displayed in the model. The advantage is that eye-movement data can be analysed automatically.

### 6.4.2 Results

Pre-study interviews indicated that operators used most of their time during the task to monitor the Fine Control Rod meter (FCR) , the effect meter, doubling time meter and the monitor (video of top of reactor). The fixation frequencies for the operators showed another picture.

The operators looked outside these instruments for more than 20 times at an average during the operation. They mostly used the FCR for monitoring the task (15 fixations an average). They had 3 fixations on the effect meter, 6 fixations on the doubling time meter; and only one fixation on the monitor (video from top of reactor).The total duration of eye-point of gaze during the operation the operators looked outside the mentioned instruments for more than 40 % of the time on an average, while gazing at the FCR in 42%, the effect meter for 3%, the double time meter 12,5%, and the monitor 2,5% during the task (see **Figure 11**).

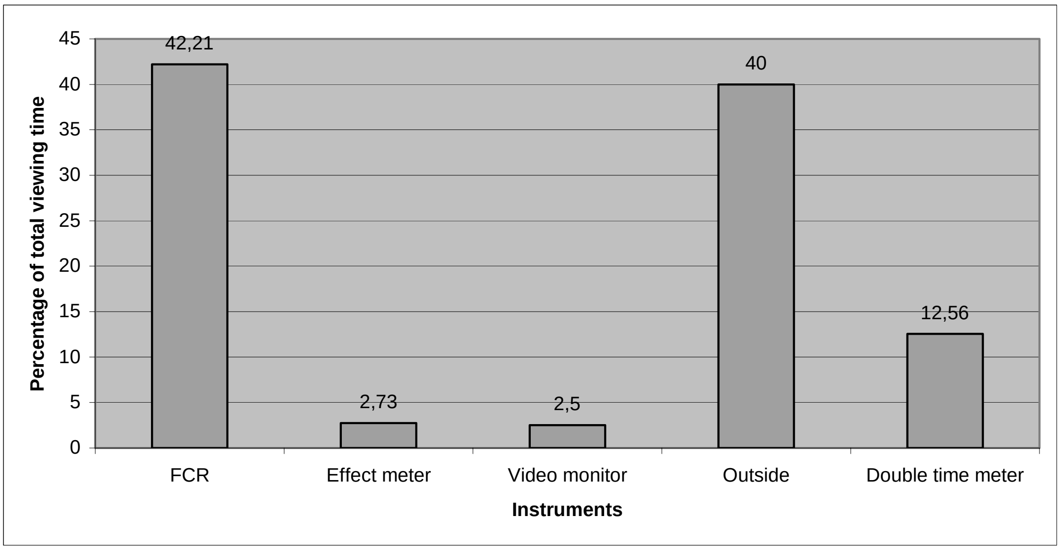



**Figure 11 Viewing time on areas of interest**

This means that although the operators in the control room had the possibility to watch the actual removal of the rod they prefer not surprisingly to monitor the task through the different meters. The data shows that they mostly use the monitor to see that the operator is at the top of the reactor, so they can tell him (through an intercom) to start the task.

***Figure 12*** shows a model of the task. The two involved operators initiate the task in the control room in co-ordinating who is to what about the size of the rod. The operator on that top (OP) prepares for the removal of the rod, while the operator in control room (OC) goes through the log book to check size of the rod, and data and time for removing the rod. While doing so he looks at the monitor a couple of times to see how OP progresses. When OP has finished preparation for removal. He positions his body in a certain way to in a non-verbal way to communicate through the video camera to OC that he is ready to pull the rod out.

The reason for this is that is, that he is not able to use the intercom from where he is standing. (to speak through the intercom the OP needs to press a button - it a simplex intercom). The OC sees that OP is ready to pull and issues the start command. While OP pulls the rod OC monitors the instruments. When OP is finished, he walks to the intercom and issues a "finish" command. Then OC update the logbook, while OC cleans up after the removal of the rod.

The SSA aspect of this task is most clear, we think, in that the OP knows that he is being monitored via the video channel. The OP also knows that the OC knows that when he (the OP) positions his body to communicate readiness for pulling the rod, this is a signal to issue the start command. This is of course a very simple task and there might be other ways of interpreting the task. Also we are of course aware of, that the simple technology available in the 40 year old reactor, in certain ways, provokes certain awareness activities. This could also be the case in more advanced systems, but then on a very different level.

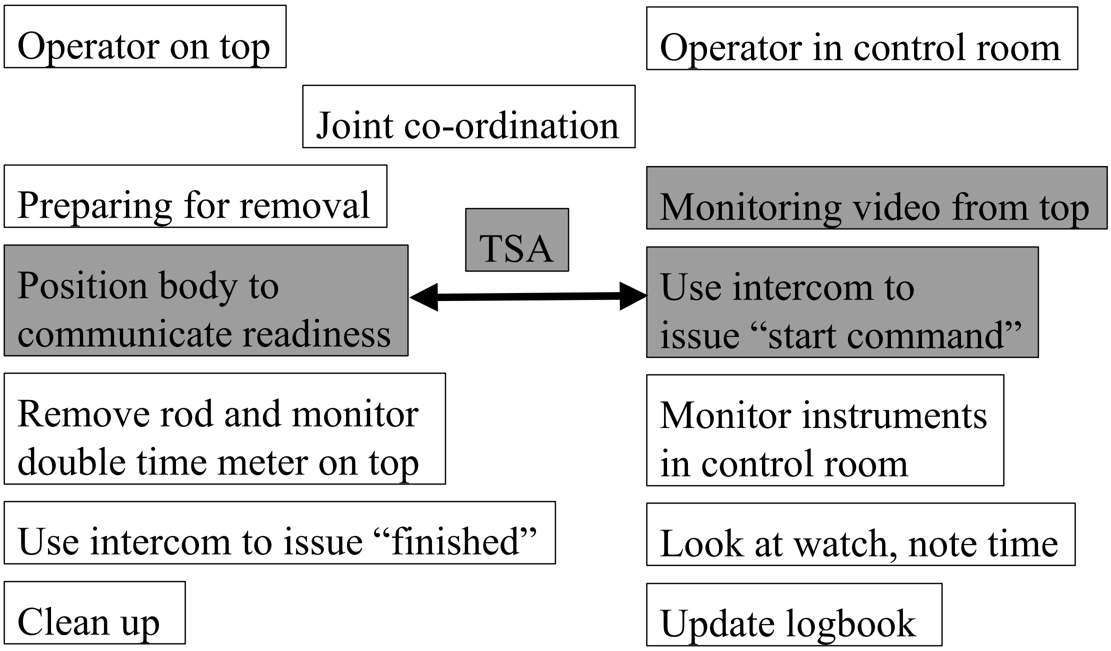



**Figure 12. A model of the task**

In the questionnaires, operators were asked to describe the success of the rod removal. They all agreed that the task could not have been performed better. All tasks were performed according to regulations and there were no shutdowns.

Operators had not worked together on this particular task before (due to summer vacation), but they had all worked together for a long time on other tasks at the reactor. This may have played a role in the communication, as no misunderstandings were produced.

It was clear that based on visual and verbal behaviour/questionnaires/recorded instruments; operators had no problems performing this task. The removal of neutron absorbing rods is a relatively simple task with respect to perception and understanding. What make it a bit difficult is that the removal has to be done manually and that this manual task must be co-ordinated with the control room.

One of the sessions seemed a bit more difficult than the rest. This session was included in the debriefing-video and discussed in plenum with all the involved personnel at the reactor. The debriefing-video consisted of a mix between our eye-movement tracking scene video recordings, the recordings of the 3D model and various location recordings. It was supported by questions regarding the level of situation awareness, but all comments were discussed:

- There were sources of information not detectable through analyses of visual information gathering and verbal communication, like e.g. listening to the elevator driving the rods.
- Operators did not look at the instruments they claimed to be looking at (both before the study and when watching the video).
- All operators agreed on what information that was important to solve the task. However, the discussion revealed that operators, even the most experienced ones, disagreed with respect to the priority of information acquisition, i.e. exactly how to access the relevant information (from what instrument, for how long etc.).

### 6.4.3 Summary of preliminary results

Based on the video-debrief, it seems difficult to establish a sequence of information acquisition if there is no strict procedure for this, i.e. defining the norm for team situation awareness may be difficult. Although individual differences in the approach to information acquisition can be observed, these differences are not necessarily more or less correct. A good approach may be to define of a lowest common denominator of information needed to solve a task, and to avoid defining details concerning qualities of the information gathering like sequence and duration, except when there can be established a clear operational definition, e.g. like following checklists. (See also Hauland 1996 on this issue).

Although operators did use instruments differently, they all relied on the same information. It was not possible from these example trials to observe variation neither in task performance nor intermediate activities like visual information gathering (type of information) and verbal communication. It seems like a (complex) measure of team situation awareness like the one proposed, is less useful for very simple tasks, tasks not likely to produce much variation.

Situation awareness is thought to be more than exceptional attention. It includes the integration of many elements in the situation, including the projection of how the situation will develop. One could ask if judging ordinal single variables like on/off, under/above calls for the type of overview we want to measure with team situation awareness. If it does, one could propose another explanation for the lack of variation: It is assumed that mental workload is tightly coupled with the concept of team situation awareness.

The relationship between workload and situation awareness is often claimed to constitute an inverse u-shaped curve: A low level of mental workload is associated with a low level of situation awareness, a medium level of mental workload is associated with a high level of situation awareness, and a very high level of mental workload is associated with a breakdown of situation awareness. Thus, if a very low level of task complexity is perceived, one would expect this to be reflected in a low level of mental workload, and consequently low situation awareness.

This explanation is probably not relevant here however, since all tasks were performed in accordance with regulations. It is more relevant, we think, to ask if this task really required team situation awareness in the way we have defined it. The difficulty in the current task – if any – is motoric (pull slowly) and co-ordination (stop pulling on command from the control room operator). It may be relevant also to ask about situation awareness in very simple task, but the complexity of the measure (resolution and number of units of analyses) must be in proportion to the tasks to be performed. Thus, the team situation awareness measure to be developed aims at measuring awareness of complex tasks, where complexity may be reflected in, e.g., number of situation elements and the type of relationship between these elements.

## 6.5 Air Traffic Control

The main objective of this study was to further develop Situation Awareness (SA) methodology, to include measures of team SA (TSA). It is also a goal to avoid problems associated with methods based on interrupting tasks. The proposed measures of TSA will be based on Air Traffic Controller (ATC) students' visual and verbal behaviour -- as well as subjective estimates from subject matter experts and the students -- during simulator training at the Danish Civil Aviation Authorities Academy, Copenhagen Airport, Kastrup. The methodological approach is both explorative and experimental.

In Denmark, Air Traffic Control (ATC) is called Flyvesikringstjenesten, (it used to be part of SLV, but is now a separate institution). The training of ATC takes place at the Danish Civil Aviation Authorities Academy. The simulator facility is located in the Tower complex at Copenhagen Airport, Kastrup. It is a simulator facility enabling realistic simulation with several types of ATC and manned aircraft in a closely interwoven role play between simulated airspace (computer) and all the aviation roles involved (supervisors, planners, radar controllers, other countries, other sectors). The so-called pseudo pilots are professional simulator pilots, usually operating all the aircraft in one sector. The subjects were all ATC-trainees finalising their training. They were observed during the last 6 weeks of training.

The approach to TSA was based on probing events defined a priori, embedded in simulated tasks (see the Targeted Acceptable Responses to Generated Events or Tasks -- TARGETs -- methodology, Dwyer et al. 1997). Probe-events are normal traffic situations (students have sufficient knowledge and skills) designed to require attention and co-ordination within and between teams. Pseudo-pilots made sure that probes occurred as designed. The probe event could be analysed in accordance with a probe checklist (normative behaviour) or at an overall level, using the probe event to define a time window for analyses (event driven sampling; analyses of, e.g., search strategies and temporal aspects of behaviour). Both TSA and system performance measures were defined in relation to the probe events (although some of the subjective estimates relate to the overall scenario). TSA measures included:

- Pilots' real time and probe specific estimates of radar and planner TSA.
- Radar and planner: correct information acquisition by means of eye movement data
- Correct verbal distribution of this information within and between teams
- Temporal aspects of visual and verbal behaviour (of radar and planner)
- Radar and planner rating each other retrospectively.

### 6.5.1 The study

There were two empirical phases of this project: the explorative phase and the experimental phase. The explorative phase sought to reveal the contents of the TSA concept in en route ATC. This phase also sought to define system performance measures in terms of the team's effective traffic handling. The experimental phase sought to validate and implement TSA measures. Validation of TSA measures will focus on how well TSA measures can predict system performance and to what extent TSA measures are sensitive to abnormal events. Eventually, the selected TSA measures will be implemented in relation to potential differences in team SA caused by the quality of the hand-over procedure. Today, there is no standard procedure for how to deliver traffic from one team to another between shifts. Thus, hand-over procedures depend on individual effort and priorities. According to controllers (interviews), this may jeopardise safety because it can affect the TSA, and possibly even affect aspects of SA throughout the shift.

**Figure 13 The radar and planner controller workspace**

#### 6.5.1.1 Objectives

(1) EXPLORATIVE: Revealing the ATC specific components of Team SA (TSA)

(2) EXPERIMENTAL: Establishing validity for the proposed TSA measures using EPOG data

The first objective (1) is the main problem in the explorative phase: what is team SA in en route ATC and how can it be observed? An additional problem in this phase is the exploration of how to observe system performance. Semi-structured interviews and observations were used in this phase. The second objective (2) is about validating the suggested TSA measures: that TSA measures can predict system performance, and to show that these TSA measures are sensitive to important changes in a situation; comparing teams with and without abnormal events. Methods were video recordings of eye movements and other types of behaviour, recording of verbal communication, real-time and retrospective subjective estimates. Validated measures (from paragraph 2) will be used to compare teams with small versus big procedure (emphasising complexity/conflicts) for hand-over between shifts.

#### 6.5.1.2 Data collection in the explorative phase

Observations were carried out in:

- Operational environments; deciding about the specific sub-domain, (i.e. an Area Control Centre (ACC) aspect of ATC).
- Regular lectures; classroom and radar simulator, learning about the ACC domain
- Selection/test specifically designed to reveal individual abilities for team work
- Evaluation meetings with instructors, observing the use of job criteria (from the existing task analysis)

- Team Resource Management Training (participating, practical course developed by EUROCONTROL)

In addition to the observation, a series of semi structured interviews (15 interviews, 1 hour each) with subject matter experts (on ATC and SA), instructors and students was carried out.

**Figure 14 Shows the radar controller wearing the EPOG bicycle helmet**

The main result of the explorative phase was

- Domain knowledge; sufficient for the various analyses
- Domain specific definition of the team SA concept and possible ways to observe it.
- Suggested system performance measures for ACC and possible ways to observe it.
- The experimental material/design/procedures (later to be tested during initial simulator training)

That is, the explorative phase resulted in a working definition of TSA and suggestions for TSA and system performance measures. So far the results from the explorative phase indicate that SA should be defined in terms of detecting available and task relevant information in a situation, and in terms of anticipating how these elements will develop. The TSA concept definition includes co-ordination of tasks and operators taking the perspectives of each other. The measures suggested during the explorative phase were both subjective estimates and objective observations of SA and performance.

### 6.5.1.3 Data collection during the eye-point of gaze simulator experiment

#### 6.5.1.3.1 Operational TSA measures

The following operational TSA measures were applied:

- Are the events detected?
- Are the events related to colleges' tasks detected?
- Taking the perspectives of colleges
- Does subject 1 detect how well subject 2 is performing?
- Does subject 2 detect how well subject 1 is performing?
- Do both subjects 1 and 2 detect how well the other is performing?
- Do both subjects 1 and 2 detect their respective perspectives on each other?
- Are the situational models of team members ahead of the current situation?
- Are the situational models of team members updated in a timely manner?

#### 6.5.1.3.2 Experimental design

The unit of analysis was the team combination of the Radar (R) and the Planner (P) students, controlling a lower sector. Professional pseudo-pilots inserted all traffic events and completed real-time estimates of ATC SA and system performance.

All students had a calibration pass (50 minutes brake and the first 5 minutes of the calibration procedure) before each pass. Calibration started, at the latest, 5 minutes before T and was completed at T + 9 minutes. Students were instructed to show up 10 minutes before T, but this was not always the case. Thus, there was minimum of 7 minutes to calibrate each of R and P. The remaining 5 minutes and 30 seconds served partly as a calibration buffer. In addition, the teams that experienced a long task "hand-over" procedure from the previous team to be released from duty could use the five minutes to follow traffic over the shoulder of instructors.

| Selected Observations: | | n = 10 | n = 11 | n = 12 | n = 12 | n = 11 |
|---|---|---|---|---|---|---|
| Sim. Time | Pilot 10th min. | Pass 1 Normal | Pass 2 Abnormal | Pass 3 Normal | Pass 4 Abnormal | Pass 5 Abnormal |
| T + 00:00 | Instructors starting all passes, handing over traffic to the student team. Student team (R + P) operating without instructor, approximately 40 minutes. | | | | | |
| T + 14:30 | Handover Complete | **SMALL** (1) procedure | **BIG** (2) procedure | **SMALL** (1) procedure | **BIG** (2) procedure | **SMALL** (1) procedure |
| T + 15:00 | Probe 1 | Unknown | Unknown | Unknown | Unknown | Unknown |
| T + 25:00 | Probe 2 | Change FL | Change FL | Change FL | Change FL | Change FL |
| T + 34:00 | *Abnormal:* | **NONE** (1) | *Fueldump* WITH (2) | **NONE** (1) | *Emerg. Dec.* **WITH** (2) | *Emerg. Dec.* **WITH** (2) |
| T + 35:00 | Probe 3 | Diversion | Diversion | Diversion | Diversion | Diversion |
| T + 45:00 | Probe 4 | Diversion | Diversion | Diversion | Diversion | Diversion |
| T + 55:00 | Clock Stop Students completing questionnaires independently, estimating each other's performance | | | | | |

**Table 12 Probe based design**

The impact of the complexity variable was explored. There can be comparison between: (a) 23 small procedure versus 12 large procedure, and (b) 12 normal versus 23 abnormal passes. Variables: event driven sampling only; probes 1 and 3. These two probes occur directly after the independent variables. A time interval has been defined around each probe event, i.e. start and stop criteria for the probe. This will be approximately 5 minutes, i.e. approximately 10 minutes video/audio tape for each recording in each observation. There were 3 video/ EPOG tracking recordings (maybe only 2) and 3 audio recordings of these 10 minutes. One-hour EPOG tracking data requires 5 hours scoring. The scoring of verbal communication has been rather demanding. All eye-track video has been analysed using the Noldus Observer video analysis software. The verbal communication has been categorised directly, i.e. without full transcription with this software package as well.

#### 6.5.1.4 Set-up in the Radar Simulator

Setup

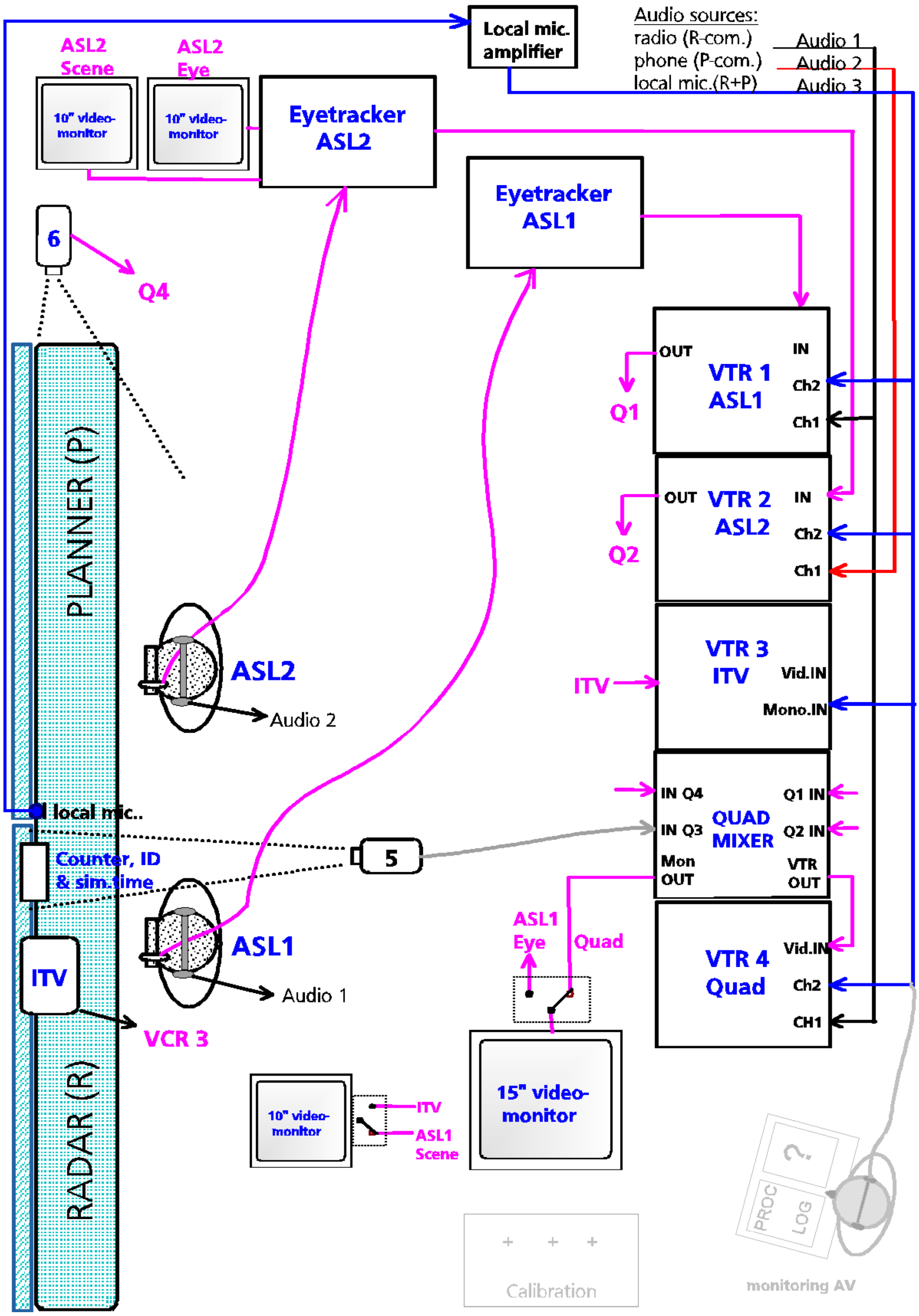



Figure 15 Set-up in the radar simulator

***Figure 15*** illustrates the experimental set-up during data collection in the full scope ACC radar simulator at Copenhagen Airport, Kastrup. The figure shows all main cable connections, all main equipment components and the actual relation between positions (although not to actual scale). Positions are the Radar and Planner positions, the external calibration position and the observer position. In ACC terms, this is the position for sector B. The other positions are locations (not to be mixed with the ATC term position). The figure shows the student team of Radar in front of the radarscope, and the Planner in front of the flight progress table.

Both operators wore eye-tracking helmets in addition to their ordinary headset with microphones. (Smaller adjustments had to be made to the headsets). There were three audio recordings: number 1 was the radio communication between pilots and radar controller. Number 2 was the telephone communication between planner and other parts of ATC. When the telephone was not in use, audio 3 listened in on audio 1, i.e. on the radio communication between pilot and radar. Note that the planner could only listen and the telephone had priority over the radio.

This set-up is part of the default ATC set-up and it enabled the planner to partly monitor what was going on, seen form the radar point of view. Audio number 3 was the local communication between the radar controller and Planner Controller. Audio 3 could usually not be heard, or heard very weakly as background, in Audio 1 and 2. Audio 3 was recorded onto all videotapes for data synchronisation and redundancy purposes.

There were 6 video in signals in this set-up, but only 4 video recordings. This was done because the Quad mix was the only receiver of two of the video signals, producing the quad video recording. This was done to reduce cost (videotapes), set-up complexity and for data synchronisation purposes. The camera targets were: EPOG tracking 1 scene with white superimposed cross hair marker, EPOG tracking 2 with black superimposed cross hair marker, camera 5: the identification image (simulator time, identification strip and transmission counter), camera 6: the position overview camera, the quad mix and camera 3: the ITV camera.

The camera numbers also identified the specific viewing angles/specific scenes used when taking still pictures and videos of the set-up and/or of the equivalent position in the real ACC (Camera 6)

Thus, there were 4 video recordings for each observation: two full size EPOG tracking recordings, one quad recording (Q1: EPOG tracking 1, Q2: E EPOG tracking 2, Q3: ID, Q4: Overview) and one ITV fill size recording (part of the log.)

#### 6.5.1.4.1 AV Recordings

**Log purposes:** The quad had the simulator time and the ID strip recorded in Q3. The ITV camera was recorded on a separate (full size) video. This recording was mono with audio 3 (for synch purposes) and it included the ID strip (for identification purposes).

**Redundancy purposes:** With 6 cameras, there was a limit to the number of backups that could be made during recording. The main data tapes ASL 1 and ASL 2 were also recorded in quad 1 and 2. The ID was recorded in the quad only, but number of transmissions was noted on the instructor score sheets and in the experimental log. (Also to avoid too much video analyses). The ID strip (file notification system) was also noted in the experimental log. The overview was recorded in the quad only.

**Synchronisation purposes:** The Quad was a simultaneous recording of the two EPOG tracking videos, performance measures, simulator time, ID strip and overview. Audio 1 and 3 were recorded on this tape. Thus, the quad was the master reference for synchronising data from both video and audio. In addition, the clap tree is recorded

**Data tapes for analyses:** The two full sizes EPOG tracking tapes with audio 1 and audio 2. These tapes were the main data recordings and they were therefore marked with a green label. The quad mix could also serve as a data tape: the performance measure number of radio transmissions was video taped in Quad number 3, the ID recording. The overview image, showing strip handling, writing, body language, etc, contained data referred to as other behaviour (than visual/verbal).

#### 6.5.1.4.2 Combined data analysis of types of observable behaviour

Basic coding of eye movements, verbal communication and other intentional actions has been merged into one data file. This means that the AV tapes have been watched coded several times, one for each type of basic data scoring.

The following data has been merged:

- EPOG: Areas of Interest (AOI a priori scoring)
- VERBAL: verbatim transcription of all ATC relevant communication, within and between.
- ACTION: scoring of Planner writing and mounting strips

The following categories of data have been collected:

- Eye point of gaze
- Radar, and planner controllers simultaneously
- Verbal and non-verbal communication
- Debriefing questionnaires
- Strips
- Number of transmissions
- Experimental log

The following categories of behaviour have been analysed:

Information acquisition
Co-ordination (content and form)
Verbal communication related to detected events
To whom, how, in what form, and at what time
Dwells on areas of interest
Radar controller: Own sector on radar, other part of radar, ITV, Planner board, the planner, other
Planner: Update of strips, radar, radar controller, other
Distribution of visual attention with respect to existing and future situations

The following shared AOI's have been coded

**Radar controller**

- ITV, the whole display
- Strips at the planner board (any designator/parallax problem)
- The planner (face/horizontal position)
- Other than defined AOI's (both a priori and empirically)
- Radar Scope for each of four regions:
    - ROSTRUP: EKRO
    - BRANDE: EKBR
    - DALMOSE: EKDA
    - STENSTRUP: EKSR
- Elsewhere Radar Scope
- Following airways
- Jumping between clusters of a/c and/or single a/c (monitor/search strateghy)
- Staying within a cluster of a/c and/or single a/c (monitor/search strategy)
- Radar Scope: The sectors C/D area

**Planner Controller:**

- Five regions:
    - ROSTRUP designator
    - ROSTRUP landing designator
    - BRANDE designator
    - DALMOSE designator
    - STENSTRUP designator
- Other designators
- The radar scope
- The radar controller (face)

In addition shifts from global to local monitoring/search and probe relevant information gathering/attention for each actor has been coded. Moreover a number of individual AOI's has been coded for each actor, e.g.,

The EPOG on AOI's has been coded in terms of

- Frequency
- Sequence

- Duration

The following categories of verbal communication have been coded:

- Number of ATC communication acts between Radar and Planner
- Number of ATC communication acts from Planer to other ATC's
- Number of social communication acts (same)
- Following probe relevant communication; yes/no scoring
- Tense of ATC relevant communication acts
- Instructions
- Requesting information
- Giving Information

The verbal communication of ATC (relevant statements made) has been coded:

- From ATC to pilot
- Between R and P
- From P to other branches of ATC

The following categories of action data have been coded

- Planner writing on strips & notes, time total writing on strips and notes
- Time total mounting strips into brackets
- Total time placing strip brackets in the correct order

The total number of observations recorded was 94. However, only 56 of these observations were selected for analyses -- based on experimental control and technical quality, as noted in the experimental log. These 56 observations constitute a time series during the three last weeks of the radar module, thus learning is down to a minimum.


**Figure 16 shows the Quad Mix of video data. Quad 1 (upper left) is the radar eye-track. Quad 2 (upper right) is the planner eye-track Quad 3 (lower left) is the ID recording: counter for number of radio transmissions from Radar controller, ID strip (data notification system) and the simulator time. Quad 4 (lower right) the overview camera.**

Raw data are AV tapes (see ***Figure 16***) from the 56 selected observations with 4 probes in each observation. From these observations are selected a (reduced) sample of 35 observations and 2 probes for each observation. If time, the remaining observations and probes will also be considered for scoring. However, the priority is to complete the TSA scoring of the reduced sample rather than having basic scoring of the whole sample. For an example of the analysis of EPOG data see Section 4.4.

### 6.5.2 The Debriefing Questionnaires

| | | | |
|---|---|---|---|
| Your pos.: R / P | Weekday: | Date: / -00 | Pass no. (1-5): |
| Instr.name hand-over: | R | P | |

Please fill in the questionnaire without others seeing you answers. "Situation awareness" means, among other things, that you "are aware of , understand ,are up-front, and can look ahead with respect to a given situation." Put a circle round your answer / mark your grade from 1 (bad) - 5 (best). 3 means medium performance. Please answer all questions. Eventually you can strike through the question that you think you cannot answer. You can put comments on the backside of the paper.

| | |
|---|---|
| Traffic situation:<br>Put circle round both 1 & 2 | **1) Not busy - a bit busy - busy**<br>2) Not complex - Medium complex - complex |
| **If R**, eventually **filtering**: NON, or; Altitude/over FL: under FL: Sector: | |

**1 (very bad) – 2 (bad) – *3(satisfying)* – 4 (good) – 5 (very good)**

- Effectiveness of traffic handling in general? 1 – 2 – *3* – 4 – 5
- Did any procedure violations occur, or where there any risk of violations?
- never - seldom - some - often - very often
- To what degree did you keep an eye on your colleges tasks (R/P)? 1 – 2 – *3* – 4 – 5
- To what degree did you keep an eye on relevant traffic outside you sector 1 – 2 – *3* – 4 – 5
- Mark your situation awareness immediately after hand-over: 1 – 2 – *3* – 4 – 5
- Mark you colleges (R/P) situation awareness immediately after hand-over: 1 – 2 – *3* – 4 – 5
- To what degree did you keep your situation awareness in the rest of the pass? 1 – 2 – *3* – 4 – 5
- To what degree did your college (R/P) keep up his/hers situation awareness in the rest of the pass?? 1 – 2 – *3* – 4 – 5
- **Mark for both you and your college if it was an abnormal/critical incident**:

Type of incident / Call sign Situation awareness immediately after the incident:

R 1 – 2 – *3* – 4 – 5

…………………… ………………………….. P 1 – 2 – *3* – 4 – 5

- What kind of effect would occur if R and P had to switch position (put circle around)

Positive, Negative, None

- To what degree do you think your college (R/P) knew about how well you performed with respect to your general degree of situation awareness?

<u>I don not know</u> OR <u>My college did not know</u> OR : 1 – 2 – *3* – 4 – 5

- Did you have any impression on the situation awareness of the other sectors? NO / YES.
  If yes : 1 – 2 – *3* – 4 – 5

## 6.6 Anaesthesia

In preceding VINTHEC II documents (WP1 Deliverable) we have proposed that Team Situation Awareness should be defined in terms of team members' *mutual knowledge.* The "Team" aspect of "Team Situational Awareness" is meant to indicate the awareness goes beyond the mere summing up of the situational awareness of the individual team members. Indeed, TSA is neither the summing up (the union) nor the coinciding or shared (the set theoretical intersection) of the team members individual awareness.

Mutual knowledge, which is distinguished from shared knowledge and common knowledge, involves higher order intentional states in the form of A's beliefs about B's beliefs about A's beliefs etc. In this section we suggest that Eye Tracking data be gathered and analysed in observational and experimental studies to indicate operators' perception of their team-mates actions and activities including their gaze.

Most technical work settings involve *teamwork* and typically require co-ordination between teams. So, humans working in crews or teams will co-ordinate their performance between themselves in order to achieve shared work oriented goals. People working together in a technical setting (cockpit, operating theatre, ATC tower, ship's bridge, etc.) will co-ordinate their response in either an explicit mode (planning, discussing action options) or implicitly by listening to and observing or just catching a glimpse of each other's activities. In this section we are mainly focused on the implicit co-ordination and its role in establishing and maintaining co-called *Team Situation Awareness*.

There are some rather distinct lines of background research that we recommend be brought to bear on TSA5. These lines of research cover

(a) the notion(s) of mutual knowledge (belief) as introduced and applied in linguistics (pragmatics - see e.g., Clark, 1986) and philosophy of language (Grice, 1957) and later applied in Artificial Intelligence and game theory.

(b) the basic notions of intentional relations including beliefs about as applied in developmental psychology and ethnology (e.g., Barresi & Moore, 1996).

Finally, we have suggested that methods for studying co-ordination in terms of gaze and gestures as well as speech derived from ethnographic studies within CSCW (computer supported co-operative work) be applied in addition to traditionally ET data analysis and that theories and results about gaze recognition and social cognition be used as inspirational sources as well.

Rather, as has been detailed in the previous VINTHEC documents about SSA, the "mutual knowledge" requirement on SSA is a compact way of summing up the constraint on team or crew members that each of them should know of each other what their colleague is attending to and not attending to.

This complements empirical studies of breakdowns of TSA (in Air Traffic Control) where incident investigation reports have revealed instances when team members (typically, a radar and planner controller) misinterpret the knowledge and awareness of their colleague. ("He did not inform her, since he thought it was routine [and he therefore thought she knew]" (Swedish CAA, ANS: S961501))

We have suggested that for Team Situation Awareness to obtain, for any given "team" or crew, the team members involved

- shall not have unnoticed conflicting awareness (so they may have different and event conflicting awareness, but they are, if TSA obtains, aware of their differences)

5 For reasons of space we cannot here list nor discuss the many relevant references to these background lines of research. However, extensive references are provided in the technical reports by the authors: see e.g., Andersen 1998; 2001; and Pedersen et al., 2001).

- have agreed on a task allocation arrangement so that cognitive resources supplement each other ("he will monitor parameters that I don't monitor, and vice versa, and we both know mutually which parameters the other is tending to").

Before seeking to characterise *team* situation awareness (TSA) in relation to the anaesthesia study let us recall that, following Endsley (1995), the individual situation awareness of an operator is often defined in terms of an operator's three-fold accomplishment of three generic tasks: (1) picking up perceptual cues of the system to be controlled; (2) integrating these cues into a coherent and valid dynamic model; and (3) predicting future states. Definitions roughly along these lines (with which we largely concur - but confer the additions to follow) will emphasise that SA consists in maintaining and updating a coherent representation of the situation that is sufficiently comprehensive and valid to allow the operator to meet current goals.

Looking at the individual operator, this type of characterisation of SA may be expanded and elaborated along the lines mentioned. But if the operator is working in a setting where he or she needs to know and make assumptions about the beliefs and awareness of other operators (say, about his or her co-pilot; or about the surgical or the anaesthetic team), the characterisation is fundamentally inadequate. For in this type of case, the "situation" of which a competent operator needs to maintain awareness will include not just the system to be controlled but also - and very importantly so - the knowledge and awareness of his / her team-mates. It hardly needs arguing that the resources required to solve the tasks will involve team-mates knowledge and awareness and priorities; nor that awareness of the "situation" cannot exclude these.

Following the suggestion by a number of authors, we distinguish between an operator's long-term and his or her short-term (situation bound) knowledge of his or her domain and work setting. Thus, Cooke et al. (2000) distinguish between what they call an operator's *mental model* and his or her *situational model.* Whichever label we use, this distinction is clearly relevant and needed. In addition, however, for teamwork to succeed, team members will have formed models of their fellow team-mates - so, we may talk about a *team-mate model.*

However, we need to apply the same distinction between a long-term and situationally determined model to an operator's conception about and expectations vis-à-vis his or her team-mate. An operator will have some generic expectations concerning the knowledge, competence, work goals, practices, norms etc. that a fellow team member will have and follow (this is part of the professional culture) even before they have met. For instance, pilots in larger airlines will meet and fly together for a few days after which they are not liable to work together for several months or even years. These generic expectations correspond to a non-individualised team-mate model and are not specific to any given day or situation. Then, for any given work session the operator will form a specific "situational team-mate model" - that is, a dynamic model of his or her fellow team member's current awareness including priorities and possibly his or her shortcomings and strengths. It goes without saying that a team member's expectations about his or her colleague's competence and norms will be individualised and much richer if they have worked together for some time.

### 6.6.1 Measuring Team Situation as Mutual Knowledge in anaesthesia

As we argued in section 2, TSA needs to be defined in terms of higher-order representations and intentional states (knowledge, beliefs, trust, goals etc.) mutually held by team members. (For details, see Andersen, 2001, and Pedersen et al., 2001). Consider the following descriptions of intentional states: (a) the nurse notices the CO2 level rising; (b) the doctor does not consider that the CO2 level is abnormal; (c) the nurse thinks the doctor has noticed that the CO2 level is abnormal, but is not sure; (d) the doctor notices nurses fidgeting with anaesthesia monitor and realises she is calling his attention to the CO2 display (sub-part of the monitor display).

Clearly, statements (a) and (b) refer to first-order intentional states, that is, a subject's state of belief (or non-belief or awareness vs. lack of awareness) , whereas (c) refers to a second-order intentional state - the nurse's belief about the doctor' beliefs. Finally, (d) may be construed as a third-order or even higher-order state: imagine, for instance, that in the scenario the doctor is extremely busy and the nurse thinks he might be annoyed if alerted in a direct way to a parameter he has already noticed: "I am afraid that, if he has already noticed the CO2, he may think that I have little trust in him [i.e., believe him to be inattentive]".

As we have shown in section 2 the CSCW literature that involved ethnographical methods contains somewhat similar observations about the achievement of team members' mutual awareness - and recognition of lack of mutual awareness - of system states. For instance, Heath and Luff (2000) point out in their well-known study of co-ordination in the London Underground Control Rooms that the "mutual availability of the various information allows personnel to presuppose that information available to one is available to all; a presupposition which is dependent upon the systematic ways in which the individuals monitor and participate in each other's actions and activities….For example, a glance towards the fixed line diagram, a gesture towards the radio phone … can …provide resources through which a colleague can recognise the actions and activities of another." (ibid., p.121).

On the classical definition of mutual knowledge, two persons, A and B mutually know that *p* if: (i) A and B both know that *p*; (ii) both know that the other knows that p; (iii) both know that (ii) obtains; (iv) etc. - up to any level of mutual knowledge.

Clearly, in complex real time domains, team members cannot share all situational knowledge but they must distribute attentional resources. Therefore, it would be misguided to try and define TSA in terms of (just) mutual knowledge of situational parameters. Our proposed characterisation of TSA is therefore (in very brief terms) that it is a necessary (not a sufficient!) condition for team-mates A and B to maintain TSA that they have mutual knowledge of each other's basic professional competence, of their *de facto* shared and *de facto* distributed tasks ("who is monitoring what and who is taking care of which task?"), of the values and significance (interpretation) of parameters within shared tasks, of situational priorities and finally, that each team-mate succeeds in maintaining SA of his or her task domain6. In the following sections we describe a proposed technique for assessing mutual awareness of system parameters and we touch briefly on results from a pilot study of operators' use of gaze and visual orientation to co-ordinate and inform each other of actions and concerns.

### 6.6.2 Gaze and Eye Movements Serving as Cues to Team-mate Attention

As alluded to above in connection with the Heath and Luff' study of the London Underground controllers, operators achieve the greater part of their co-ordination through implicit means. They rely on shared visual and auditive cues in shared work space (shared in the sense that they are - and are known by team members to be - readily available to both team members); they pick up their fellow team members' direction of attention (or lack of directed attention) by noticing gaze direction and they let each other know what they are attending to by direction of head and of gaze.

The ability to shift our attention in the direction towards which another person’s eyes is turned seems to be an innate competence. For instance, Hood et al. (1998) report that "infants as young as 3 months attend in the same direction as the eyes of a digitised adult face". Similarly, Langton et al. (2000) note that the structure of the eyes of humans "provides us with a particularly powerful signal to the direction of another person's gaze", and they point out that gaze direction "is analysed rapidly and automatically, and is able to trigger reflexive shifts of an observer's visual attention" (ibid., p. 50).

### 6.6.3 The study

The authors conducted an observational pilot study of anaesthetists' performance and visual behaviours during critical patient scenarios in a comprehensive anaesthesia simulator.7 The pilot study was made in

6 Several additional conditions - notably integration of information - need to be added to characterise TSA. See Andersen, 2001 for further discussion.

7 We are grateful for support and generous advise from the Danish Institute of Medical Simulation at the Herlev University Hospital (Copenhagen). Our special thanks go to Anne Lippert, MD, Dept of Anesthesiology, Herlev University Hospital.

order to (1) test the usability and face validity of the Assessment of Team Situation Awareness questionnaire (see below) and (2) collect and analyse EPOG data from realistic, critical scenarios.

**Figure 17 Anaesthesia doctors during a simulation in the simulator at Herlev University Hospital in Copenhagen**

But the study was not a clinical trial, there were no (in)dependent /dependent variables, and most importantly, the number of trials (i.e., 2) and subjects (i.e., 2+2) was too small to permit generalisations and inferences)8. Yet, its results may serve to illustrate the use of ET data to reveal co-ordination mechanisms and phases at which TSA breaks down. During each scenario, the doctor and the nurse wore eye tracker helmets, and recordings of their visual behaviours were supplemented with one video recording of overall operation scenery and one video track of the monitoring screen.

### 6.6.4 Results

The study revealed that the doctors and nurses, while rarely focusing on each other's gaze, at crucial points seemed to follow each other's line of gaze. There was a preponderance of simultaneous dwells on areas of interest (patient, monitor) though a consistent division of labour was also observed. In addition, their chief implicit mode of acquiring awareness of their team-mate's activities was simply to visually sample what the hands of their team-mate was doing. During one very hectic episode of one of the scenarios when the doctor was busy and highly concentrated administering IV infusion while verbalising his worries and hypotheses to the surgeon, the nurse urgently needed confirmation that the drug she had in her hands was the intended one. Therefore, she waved the drug label in front of the doctor's field of vision and he nodded. Finally, subjects were observed to perform an additional visual check on the monitoring apparatus whenever their team-mate announced a slightly deviant or unexpected values.

8 Two sessions were conducted, the sessions being different both in terms of the scenario (script) used and trainee team. Each team consisted of a physician in training (2$^{rd}$ or 4$^{th}$ year of specialising) and an experienced nurse (5 or 8 years of anaesthesia experience). Each session lasted 35-45 minutes. Both scenarios required tight team collaboration when the critical symptoms were introduces (one scenario involved a surgeon-induced vein puncture, the other was involved a severe allergic reaction approaching an anaphylactic shock).

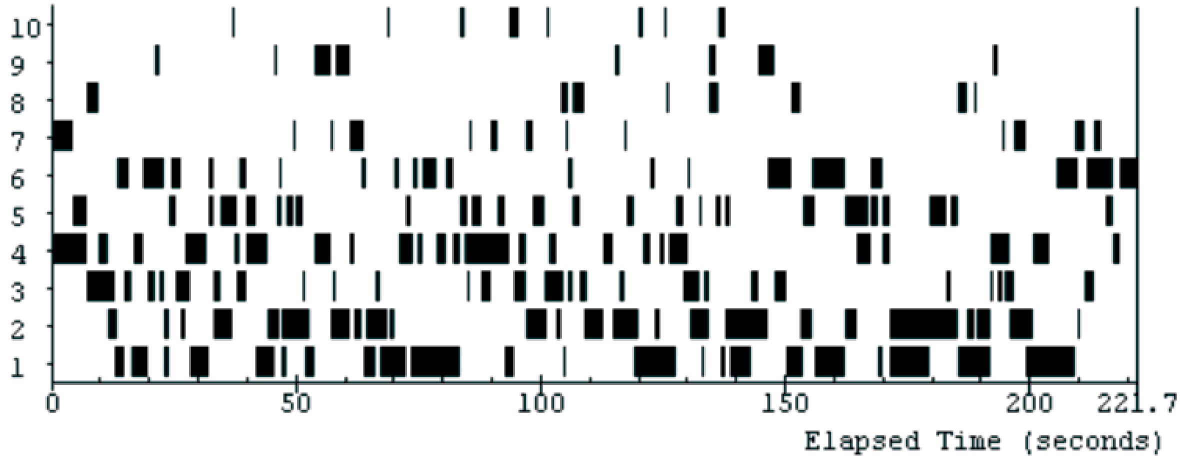



**Figure 18 Time Event plot team 1 doctor and nurse dwell on areas of interest Legends: 1. Doctor gaze @ monitor 2. Nurse gaze @ monitor. 3. Doctor gaze @ patient. 4. Nurse gaze @ patient. 5. Doctor gaze @ instruments. 6. Nurse gaze @ instruments. 7. Doctor gaze @ nurse. 8. Nurse gaze @ doctor. 9. Doctor gaze @ surgeon. 10. Nurse gaze @ surgeon.**

The Time-Event Plot method produces a graph in which observational data are plotted against a time axis, in other words, a time-event plot is a time-event table in a graphical layout. The EPOG on defined AOI's are plotted horizontally against elapsed time. A time-event plot can be used to get a first impression of the nature of EPOG data, e.g.:

- Whether dwells on defined AOI's are rhythmic or irregularly spread over the observation time.
- The variation in the duration of dwells on defined AOI's.
- The relationship between dwells on different AOI's.
- The relationship between dwells on AOI's of different subjects.
- The sequential relationship dwells on AOI's.

The degree of overlap of dwells on AOI's could be used as a TSA measure. In the anaesthesia study, it was quite clear that there is a relatively high degree of overlap between nurses and doctors. This could indicate low TSA - that the doctor checked the work of the nurse repeatedly and vice versa. On the other hand, it could also indicate the formalised division of labour between the doctor and the nurse - that the doctor has the full responsibility for team performance.

### 6.6.5 The ATSA debriefing questionnaire

Based on the above conception of the cognitive co-ordination involved in real time team work, we have devised a short battery of questions (ATSA: Assessment of Team Situation Awareness) designed to elicit a subjects' estimates of his or her team-mate's situation awareness and view of task allocation in addition to the subjects' own first-order knowledge of significant system parameters and for the VINTHEC II project to capture these basic SA and TSA measures. The brief questionnaire is intended for use (and has been applied in pilot studies of anaesthesia simulations involving medical Crew [Team]Resource Management training) during pre-planned interruptions of audio/video prompted *debriefing* sessions immediately following an experimental trial or a training session (conf. Hansen, 1991, for descriptions of this debriefing technique). The ATSA questionnaire *may* be used in the intrusive and interruptive manner the original

SAGAT was used (interrupting a simulation session and eliciting subjects' estimates, continuing the session and then conducting a repeat interruption and elicitation of awareness judgements). However, the *intended use* of the ATSA is that it should be applied during post-session; and indeed, this is the manner in which the ATSA has been prototype tested during the small-scale experiments of the WP2 phase.

The ATSA questionnaire taps individual crew members' (a) awareness of selected system parameters, (b) estimate of their fellow crew member's awareness (correctness in assessing a given parameter), and (c) perception of allocation of monitoring responsibility. At the same time, we have asked crew members to assess their own and their team member's workload.

When comparing responses from crews we shall be able to gauge

(a) the accuracy of individual crew members' estimates of system parameters (and, by extension, the agreement between crew members)
(b) the ability of crew members to correctly predict the awareness of their fellow crew member
(c) the extent to which crew members may correctly predict the workload of their colleague and their colleague's perception of task allocation.

In the following table we have summed up the measures of the ATSA questionnaire divided into individual SA and team SA measures.

**Table 13 ATSA debriefing questionnaire**

| **Assessment of Team Situation Awareness - table of measures elicited** | | |
|---|---|---|
| **Measure** | **Individual SA measure:** Respondent's individual awareness | **Team Measure: Awareness of partner's awareness (TSA)** |
| workload | 5 point scale | 5 point scale |
| values of system parameters at time of interrupt | for each parameter, current | for each parameter, is my partner right / reasonably right / possibly far off |
| trend of system parameters (within last couple of minutes) | indicate whether parameter has been rising, falling or been stable | **nil** |
| your confidence in your own parameter estimate | 10 point scale | **nil** |
| who is responsible for monitoring this parameter during ***this*** phase | indicate whether responsibility for parameter monitoring is shared, respondent's or colleague's | indicate whether colleague will agree with task allocation (parameter monitoring responsibility) |

The ATSA debriefing questionnaire was demonstrated tested during the small scale experiments in two anesthesia simulation sessions. 9

Two sessions were conducted, the sessions being different both in terms of the scenario (script) used and trainee team. Each team consisted of a physician in training (2rd or 4th year of specialising) and an experienced nurse (5 or 8 years of anaesthesia experience). Each session lasted 35-45 minutes. For both scenarios tight team collaboration was essential when the symptoms were introduces (one scenario involved a simulated surgeon induced vein puncture, the other was a allergic reaction approaching an anaphylactic shock).

During each scenario, the doctor and the nurse wore eye tracker helmets, and recordings of their visual behaviours were supplemented with one video recording of overall operation scenery and one video track of the monitoring screen.

The ATSA questionnaire was applied during debriefing. During the debriefing sessions each of the 2 x 2 subjects were confronted with the video recordings of their own visual behaviors (and the three simultaneous recordings of the global scenery, the visual orientation of their colleague and the monitoring screen). Each subject was introduced to the ATSA questionnaire and a familiarization trial (filling out of the questionnaire) was made for a point in time for the normal pre-operative phase. Then, For each session 3 or 4 pre-planned interrupts were made. During the interrupt (about 4-6 minutes) the subject would fill out the ATSA questionnaire.

The experience with the questionnaire included the following points:

-subjects seemed to "re-live" the scenario with no difficulty

-while no attempt at all was made to counteract the hindsight perspective ("at this time I thought the CO2 level was normal, but of course it was not"), subjects themselves volunteered on several occasions their in-session erroneous perceptions of system parameters

The most difficult aspect of analyzing data from the questionnaire lies in defining the appropriate intervals, for each parameter, that define when an estimate is "right", is "approximately right" (i.e., the estimate is not quite right, but it is not critically wrong) and when "critically wrong". It turned out that it would be exceedingly complex (and possibly too complex to process and validate) to define in quantitative terms, across any scenario, across all possible patients, these three intervals; in contrast, it turned out to be reasonably straight forward for an expert to define these intervals for a specific scenario, a specific patient and a specific phase.

9 The experimental task environment is an operating room (university hospital, Herlev Hospital) in which a comprehensive anesthesia simulator is operated. The anesthesia simulator is used for conducting training sessions for anesthetist nurses and physicians

### 6.6.6 Assessment of Team Situation Awareness (ATSA) Debriefing Form

For each interruption during video review each trainee shall fill out a copy of this form

| Subject name / no.: | Session no.: | Team no: | Interruption no.: |
|---|---|---|---|

| Please estimate your and your colleague's workload **right now**? | **Very high** (am using all my attention on my tasks; will refuse interruption by external calls) | **Rather high** (have few resources to spare; would be reluctant to accept external calls) | **Medium** (am occupied but do not feel any great load; will accept simple external calls) | **Rather low** (am slightly occupied; external calls would be welcome) | **Very low** (merely monitoring a normal, non-complicated anaesthesia; external calls welcome) |
|---|---|---|---|---|---|
| Your own workload - insert tick mark → | | | | | |
| Your colleague's workload - insert tick mark → | | | | | |

| | | Pulse | Systolic | Diastolic | Oxyg. sat. | CO2 |
|---|---|---|---|---|---|---|
| **Your own estimate of parameters** | Current trend - within the last couple of minutes the parameter has been rising, level or falling (please insert ↑, →, or ↓ ): | | | | | |
| | Write your estimate of the value of the parameter: | | | | | |
| | Indicate your confidence in your previous estimate on a scale from 1 to 10 (1 = entirely uncertain, 10 = entirely certain) | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Your estimate of your colleague's know-ledge of parameters** | Please enter **Y**(es) or **N**(o) or "**---**" (= don't know) depending on whether you think your colleague will provide a correct or wrong estimate of parameters | | | | | |
| **Task allocation** | *Whose* task is it to monitor the parameter in this phase? Please enter **M**, **B** or **C** ("primarily **m**ine", "shared=**b**oth of us" "primarily my **c**olleague's"): | | | | | |
| | Do you believe your colleague will agree with your view of task distribution?<br>Write **Y**(es) or **N**(o) or "**---**" (= don't know) | | | | | |

# 7 JAZZ and attention modes

## 7.1 Design concept

We developed the new approach to the EPOG measurement where its two basic functions, are the acquisition of the eye position and inferences about its temporal changes, are split and acquired by the two separate systems. The first system we have named „WHERE" and the other one „WHEN".

The pilot's gaze monitoring system becomes synchronised with the visual information acquisition, which is ordered by the pilot's higher function. This system we will call later **the saccadic sampling EPOG**. The main advantage of such task separation is that it simplifies the system set-up and calibration. We estimate that the time required should be not more then 5 minutes. All operations should be performed semi-automatically with only minimal human operator attendance. The other advantage is the possibility of using for each of the tasks the best measurement methods. The visual information acquisition takes place during the initial phase of the fixation. Due to the average fixation duration (about 220 ms) the number of „scene-shots" usually ranges from 4 to 5 per second,10 thus the EPOG sampling rate can be reduced to 5 Hz. The reduction of the information incoming to the system is another advantage of the "saccadic sampling EPOG" (comparing to 25 Hz sampling of the standard video-oculography signal).

It is assumed that the human visual system is much like taking the picture of the selected region of the visual surrounding. Directing the camera on the selected AOI is carried out by the saccadic system. Voluntary saccades direct the camera according to the request of the conscious brain. Two mechanisms control the release of the shutter. Decision about performing the saccade is equivalent to rewinding the film and unlocking the shutter. The retinal blur, which accompanies the saccadic movement, is involved in "controlling of the shutter". Release of the shutter takes place when the saccadic blur ceases.
This means that at the instance of saccadic landing, when the image on the retina becomes steady, the image is transferred to the buffer (iconic memory) for temporally storage. The acquisition time of the visual input is equivalent to the opening of the camera shutter. Its duration is of the order of milliseconds (one can assume that at the very low level of illumination it will take possibly more than 10 ms). The regular fixation time is equal to 220 ms in average. The question may be raised why the fixations are so long, while the acquisition of the visual information takes place during such a short period, and why necessarily at the beginning of the fixation? We expect that the rest of the fixation time is needed for two purposes.

First of all the part of the oculomotor system responsible for the stabilisation of the image on the retina, requires this time for performing the measurement of the image displacement velocity on the retina. The optokinesis and smooth pursuit systems use this velocity as the settings for the next fixation slow eye movements. The second reason is that the processing of visual information takes time.

The processed information cannot change on the fly, while the information is analysed, because it will disturb this process. This means that visual information needs to be acquired as early as possible during the fixation and should be not allowed to change during the rest of fixation. Only the next saccade can overwrite the information acquired during the previous fixation. The short acquisition time is necessary for achieving good picture quality. The retina is very sensitive to the instabilities of the projection. Displacement of the image causes the retinal smear equivalent to the blurred pictures taken with the shutter opening longer than 1/100 s, while the photographed person was for example running. Short acquisition time, as well as the precise synchronisation with the

10 If the saccade appears after the fixation shorter than 160 ms it will not result in EPOG sampling (Zuber criterion for visual information processing).

saccadic landings, requires that the system responsible for the „WHEN" part of measurement should have high temporal resolution i.e. the sampling rate at least 1 kHz.

## 7.2 System "WHEN"

System "WHEN" is the high sampling rates eye-movement monitoring system, which controls information intake. For this part **direct IR oculography** is used, combined with the video-oculography (system WHERE) into the one hybrid EM measuring system. The EM signal acquired by the IR oculography is analysed with the aim to detect the instances of saccadic landings. At such instances the video-oculography frames are acquired along with frames from the scene camera attached to the subject's head. The direct IR oculography is an effective method for the high temporal resolution measurements, but is very weak in answering the question „WHERE". Usually it will require a complicated calibration procedure, what makes the whole system not applicable in the cockpit. From the IR oculography signal one can additionally filter out the information about blink rate, which is another useful indicator of the pilot's mental workload.

The IR oculography system is also equipped with the leaking eye velocity integrator. This part of the system allows monitoring of the overall eye movement activity including blinks. Leaking time constant is equal to 3 seconds and refers to the duration of the consciousness-sampling period (duration of the Pöppel's "now"). The overall eye movement activity provides information about current pilot's mode of mental activity (exploration, monitoring or planning). For further details see the supplemental documentation for VINTHEC II project, article "Discrepancy between what they see and what they say" by J.K. Ober and J.J. Ober, 1 June 2000.

## 7.3 System "WHERE"

The system "WHERE" should provide the information about the eye point of gaze held during the fixation time. For this task, we will use the video-oculography system combined with the scene camera. The task can be divided into two parts: estimating **the eye-in-head position** (EH) and estimating **the head-in-cockpit position** (HC). Both positions will be expressed in pixel co-ordinates of the eye and scene images and connected by means of the 2-point calibration procedure. It will require the pilot to look at two targets (left and right) displayed by the laser projectors attached to the scene camera. The measured distances between those two calibration points for scene and the eye image, will provide the factor for linear interpolation of **the eye-in-cockpit position** — EPOG.

## 7.4 The eye-in-head position

The camera mounted on the pilot's head just above his eye-cavity acquires the eye image. The eye is illuminated with infrared light. The infrared filter is mounted on the front of the camera CCD sensor to eliminate the daylight influences.
Each "eye-shot" is pre-processed using the following scheme:

- The illumination equalisation — the image can be unevenly illuminated due to the asymmetry of the eye-cavity.
- Approximations of the pupil location — the darkest points of the image are the good approximation of the pupil's position. It significantly reduces the area of image considered by the subsequent processing steps.
- The image thresholding and noise reduction (the pupil area delimitation).

The binary image with roughly estimated pupil position is the start point for the 3-phase iterative pupil's area approximation using the rectangle.11 In the first phase the rectangle is moved to the position, which covers the maximum number of black points (i.e. points that belong to the pupil). The second phase implies the modification of the rectangle size with 4-pixel step. Finally (third phase) the rectangle dimensions are modified with 1-pixel step. The pixel co-ordinates of the resultant rectangle centre are taken as the current eye-in-head position.

---

11 Approximation of the pupil using the rectangle was proposed by Michał Młodkowski, IBIB-PAN — annual internal report (1998/99), Department of Biomedical Information Processing Methods

The rectangular approximation of the pupil is to some extent insensitive to the glints that occur on the eye image. Only when a glint is located on the edge of the pupil it will disturb the measurement. As the solution to this problem we proposed the acquisition of two frames, each performed with different illumination using two independent infrared LED's. In case of detection of the pupils edge disturbance the other frame can be used.

### 7.5 The head-in-cockpit position

The control of pilot's activity in the sense of eye point of gaze requires feedback information about the instruments the pilot is looking at. The scene camera mounted on the pilot's forehead acquires the part of the cockpit scenery covered by the pilot's field of view. The analysis of the "eye-shot" gives the cross-mark on the "scene-shot", which denotes the eye point of gaze. This is quite acceptable way of presenting the EPOG data when considering several images. However, during the flight many thousands of frames will be acquired (about 15000 per hour) what makes the operator's viewing of each frame impossible. Thus, the automatic eye-in-cockpit position detection should be implemented and only symbolic information should be stored (for example the numbers denoting the group of flight instruments).

Within the cockpit there are some particular areas, which contain the displays of information crucial for performing the flying task like PFD, NAV, MCP, A/P, engine display, CDU and outside view. We suggest to place near some selected instruments, simple rectangular markers made from the reflective material (3M Scotchlite™). The light from the infrared LED's mounted near the optical axis of the scene camera is reflected from the markers and returned back to the scene camera. The infrared filter situated in front of the camera's lens filters out the daylight and increases the contrast between the markers and the background. The markers should be placed in the cockpit in such a way that in every direction of the pilot's view at least one marker should be visible. Each marker will be labeled with the black gap, located at different positions along the longer axis, what will univocally identify its position in the cockpit.

For each eye fixation two frames will be acquired from the scene camera: one with infrared LED's ON and the other with OFF. The difference between these two frames will result in an almost dark image with a clear, bright marker. Such an image is the good departure point for further analysis performed in the similar way to the one used for the pupil position finding.

# 8 JAZZ – Mark1

System JAZZ-Mark1 is the advanced version of the eye movement input device to be used later in the pilots' visual attention monitor (cross-biofeedback). RISØ is currently evaluating the simple prototype version.
The photoplethysmography signal that is sensed at the forehead skin mimics the Peripheral Arterial Tonometry developed by the Technion Group (Peretz Lavie PhD et al). It is intended that later recordings will simultaneously collect both JAZZ and PAT data and to assess the degree to which these two correlate.

### 8.1 System structure

The system consists of four elements: eye movement sensor board, signal conversion board, data acquisition and control unit and retransmission device. Eye movement sensor board and signal conversion board contains a set of various sensors measuring the biological and environmental parameters: eye movement, ambient sound, pletysmography, oxymetry, two-axis acceleration and ambient illumination.

### 8.2 Eye movement

Eye movement is measured as the average of left and right eyes separately along the horizontal and vertical axis. The horizontal measurement range is ±30°, the vertical ±20°. Effective sampling rate is 1kHz (8 kHz hardware over sampling).

## 8.3 Signal conversion

Signals from the sensors are processed and converted to digital format within the signal conversion board. Converted data is later sent to the data acquisition and control device via digital link, which also provides control signals and power for the data conversion board.

## 8.4 . Data acquisition and control

Data acquisition and control unit collects the data from measurement subsystems and provides a set of control signals. Downloaded data is preprocessed using Scenix RISC micro controller unit. Data acquisition and control device can be hardware configured to be a freestanding data acquisition device. It provides +3V power supply for measurement subsystem and eight general input/output ports. Each of the pins can be configured to be either digital input, digital output, analogue input or analogue output. Number of the ports which alternatively can be used as analogue output lines is limited to 4, however, there's no limitation for other port functions.

Data acquisition devices are stackable to provide the possibility of future system expansion. Data is sent in the stack using IrDA1.1 - 1Mbit/s optical link. Single data acquisition and control device can provide the IrDA1.0 standard 115kbit/s link directly to a computer.

## 8.5 . Retransmission

Retransmission device provides the data output for data acquisition device stack. It collects data from the stacked devices and sends it through RS232 serial link to a computer. Future versions of the retransmission device will provide a USB link to a computer.

### 8.5.1 Structure of the JAZZ-Mark1 eye movement measurement system

#### 8.5.1.1 The concept

Saccadic intensity signal, defined as the eye movement velocity above the saccadic velocity threshold accumulated over 3s time windows is used to indicate the engagement of the conscious attention, (the conscious brain – cortex) in the internal processing domain or in the external one. During the brain involvement in the external domain, which takes place when exploring and monitoring (only partially) the visual environment, the saccadic intensity has its maximum. During planning – thinking about past and future as well as when logically evaluating the already internally available information, the saccadic intensity has its minimum. Saccadic system nearly ceases its activity. During such periods there is no visual input that is, the visual inattention.

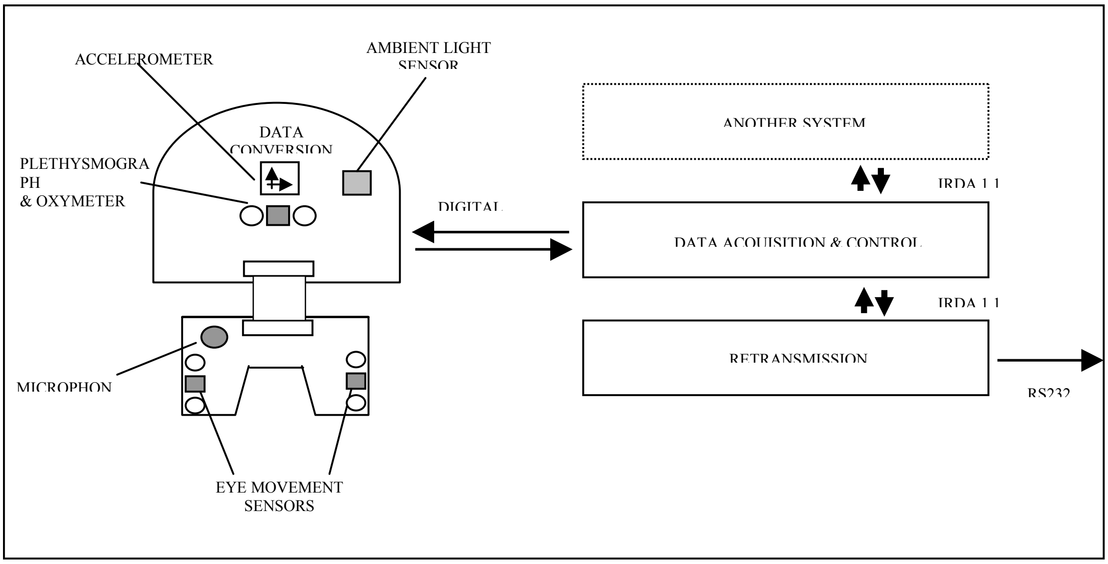


**Figure 19 The JaZZ mark1 concept**

#### 8.5.1.2 The system

The system comprises two components; the eye movement sensor and the signal processing and indication unit. The system was custom designed by the IBIB – Poland as the demonstration system allowing demonstration of the above stated concept with the purpose of conducting the feasibility study to test its applicability in the VINTHEC II project.

The system allows the measurement of horizontal eye movements, over the range ±45° with the resolution of 0.1 degree. The output voltage can be available (optional) externally. Its is changing between 0-3V with the sensitivity 10 mV per one degree. The noise level in the signal is equal 1 mV. The eye movement sensor, measures the mathematical average of the movements of both eyes along the horizontal axis.

#### 8.5.1.3 Restrictions

The cyclop sensor and its signal processing circuitry is the joint intellectual property of Ober Consulting Poland and Advanced Ballance System Ins. Columbus Ohio (USA Pat Pend.).
The cyclop sensor was adopted for visual attention monitoring and as such was developed and prototyped by Ober Consulting Poland, acting as the subcontractor for IBIB-PAN in the VINTHEC II project.

The use of the visual attention monitor is restricted for the evaluation purpose only for the VINTHEC II project. All other possible applications of its methodology requires written permission from Ober Consulting Poland.

### 8.5.2 Examples of eye image

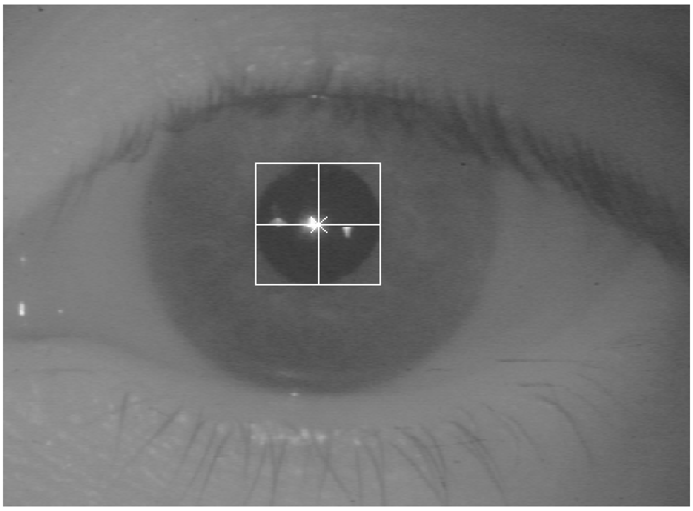


Image 1. The “eye-shot” with the pupil described using the rectangle

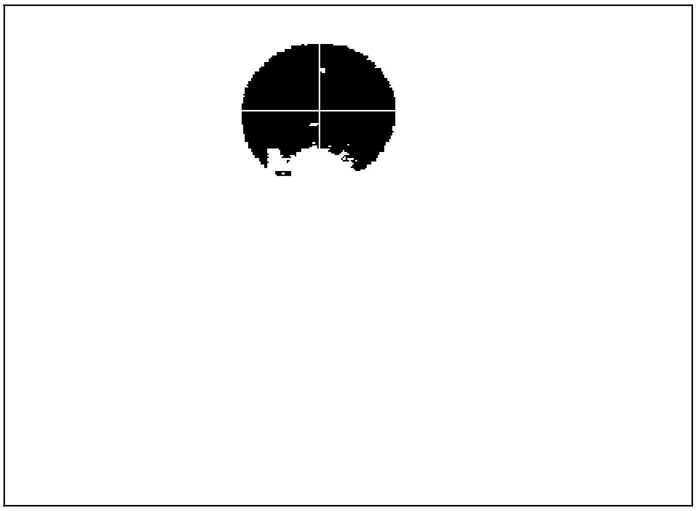


Image 2. The thresholded image of the eye — pupil edge disturbed by the glint

### 8.5.3 Examples of landmarks

Image 3. The original photo of the marker with the IR illumination

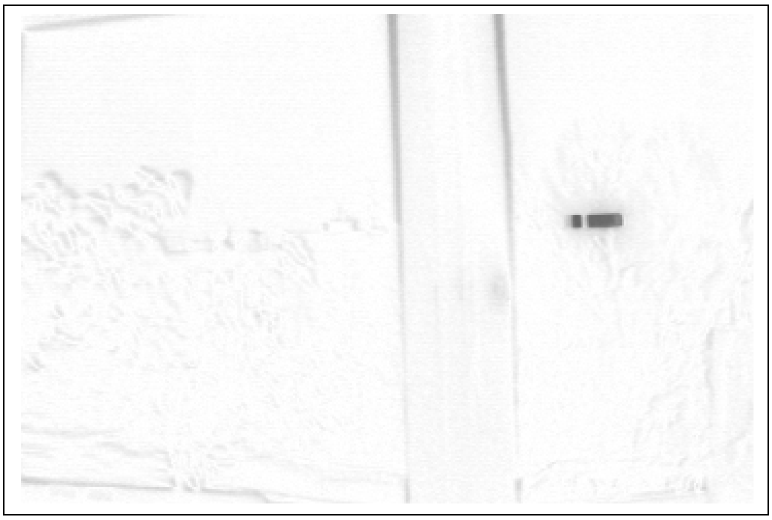


Image 4. The difference image (with and without IR illumination)

# 9 Summary of proposed methods

## 9.1 Verbal data

Tracing the information gathering of subjects seems incapable by itself of indicating how subjects use the detected information, i.e. EPOG data can not represent so-called higher cognitive processes without an analysis of a given context. For this analysis (non)verbal data are needed. We suggest that these data be derived from observations and transcriptions of cockpit voice recordings, possibly augmented with video recordings of non-verbal communication.

As a point of departure, the basic techniques of for verbal data measurement could be used (see also ***Table 14***). These comprise the transcripts of each crew-members speech, coded into pre-defined types. These codings can be analysed using three communication parameters: Number of speech acts (few/many), homogeneous communication patterns (similar/different), and temporal aspect of communication (past, present or future tense). See also section 2.5.3.

| | **Dimensions of communication** | | |
|---|---|---|---|
| | Number of speech acts | Homogeneous communication patterns | Temporal aspect of communication |
| **Types of speech acts** | | | |
| Command | | | |

| | | | |
|---|---|---|---|
| Observation (about flight or system status) | | | |
| Inquiry or request for information | | | |
| Response uncertainty | | | |
| Agreement | | | |
| Acknowledgement | | | |
| Repetitions (of already stated commands or inquiries). | | | |
| Social | | | |

**Table 14 Verbal data measurements**

## 9.2 Observational data and behavioural markers

As we see it there are four main interaction categories or modes of interaction (see ***Table 15***).

**Maintaining reciprocal awareness:** The team could be involved in synchronous activities, by monitoring colleges' location, and to monitor their activities. Moreover, they could be engaged in explicitly making their own activities publicly visible to team-mates by thinking aloud, humming, etc.
**Directing attention:** Actors attract the attention of team-mates to focus on certain features or emerging problems in the field of work by, for example, to position certain items in certain ways, by pointing or nodding at particular items.
**Assigning tasks:** Actors could for example allocate a task by nodding at a work object or by stating a verbal request.
**Handing over responsibility** of processes in the field of work, for example, by passing on the work object in question, or the interface of a control mechanism.

| | **Dimensions of modes of interaction** | | |
|---|---|---|---|
| | Unobtrusive versus obtrusive | Embedded versus symbolic | Ephemeral versus persistent |
| **Modes of interaction** | | | |
| Maintaining reciprocal awareness | | | |
| Directing attention | | | |
| Assigning tasks | | | |
| Handing over responsibility | | | |

**Table 15 Observational data and behavioural markers**

These modes of interaction are combined and meshed dynamically and fluently to meet the requirements of a specific situation. The different modes of interaction cannot be ordered in any simple kind of way but is possible to point at a limited number of prominent dimensions of the modes of interaction (see ***Table 15***). Some examples:

**Unobtrusive versus obtrusive**, that is, some modes of interaction can be disruptive in nature in relation to colleges' line of work, while others are very conspicuous and therefore permit colleges to carry on work.
**Embedded versus symbolic**, that is, to embed cues in highlighting certain items belonging to the field of work by for example marking them versus using a symbolic representation of the cues which through its abstract function offers a higher degree of freedom regarding the manipulation of the cues**.**

**Ephemeral versus persistent**, that is, shared situational awareness only appears during the course of work and then disappears without leaving any trail to track. It is for example not immediately possible to trace activities like monitoring co-workers activities or to make ones own activities publicly visible.

(See also section 2.4)

## 9.3 EPOG Measures

VINTHEC I showed that there were some individual differences in fixations patterns among the pilots.
The problem could be that if the information acquisition is associated with one fixation only, the fixation definition will also depend on context variables, and consequently vary both within subject and tasks and between subjects and tasks. Any fixation definition must take into consideration that the analysed fixation pattern depends on the definition of one single fixation, i.e. the smallest cluster of sampling points and the sampling duration. Comparing fixation definitions that vary in milliseconds can give qualitatively different fixation patterns, or scan paths.

### 9.3.1 Dwell time analysis

By definition, a "dwell" requires an area, i.e. an interface parameter, upon which the line-of-gaze is directed. A dwell consists of several single fixations, the target is an area rather than the exact location of an EPOG marker Dwells are conceptually defined here as several single fixations within an area of interest (AOI). The operational definition is that the eye movement cross hair must stop for a minimum of 200 milliseconds inside the border of an AOI (or be within 0.5 degrees visual angle from the AOI centre, to allow for small calibration inaccuracies). As soon as the cross hair leaves the border of an area of interest, the dwell is terminated. We suggest that we analyse EPOG dwells on AOI's in terms of frequency, sequence, duration, transitions. In addition we should calculate pupil diameter, blink rate, and blink duration (See also sections 38 and 4.1)

### 9.3.2 Analysing SSA based on a Time-Event Plot and Table Method

The Time-Event Plot method produces a graph in which observational data are plotted against a time axis, in other words, a time-event plot is a time-event table in a graphical layout. The EPOG on defined AOI's are plotted horizontally against elapsed time. A time-event plot can be used to get a first impression of the nature of EPOG data, e.g.:

- Whether dwells on defined AOI's are rhythmic or irregularly spread over the observation time.
- The variation in the duration of dwells on defined AOI's.
- The relationship between dwells on different AOI's.
- The relationship between dwells on AOI's of different subjects.
- The sequential relationship dwells on AOI's.

The degree of overlap of dwells on AOI's could be used as a SSA measure. In the anaesthesia study it was quite clear that there is a relatively high degree of overlap between nurses and doctors. This could indicate low SSA. - that the doctor had to check the work of the nurse over and over again and vice versa. On the other hand, it could also indicate the formalised division of labour between the doctor and the nurse, that the doctor has the full responsibility for team performance. This means that we as a basis for applying this method need an analysis of the pilots' typical tasks during the simulated flight scenario. See section 6.6.4 for an example of the use of time event plots. (See also section 6.6.4)

### 9.3.3 Visualising SSA Based on Lag Sequential Analysis

Sequential analysis is a collection of techniques developed for the study of the temporal structure of sequences of events. A commonly used method is lag sequential analysis, which allows you to calculate frequencies of transitions between dwells on AOI's within a certain lag in a time series. Lag sequential analysis allows you to answer questions like: "How many times is the a dwell on the primary flight display followed by dwell on co-pilots hands?"

The outcome of a lag sequential analysis is a transition matrix as the one shown in ***Table 16***

| | | | | | |
|---|---|---|---|---|---|
| SubjectA_behaviour1 | 0 | 3 | 11 | 12 | 6 |
| SubjectB_behaviour1 | 4 | 0 | 4 | 11 | 7 |
| SubjectA_behaviour2 | 7 | 9 | 0 | 1 | 14 |
| SubjectB_behaviour2 | 1 | 12 | 17 | 0 | 2 |
| SubjectA_behaviour3 | 5 | 5 | 7 | 8 | 0 |
| | SubjectA_behaviour1 | SubjectB_behaviour1 | SubjectA_behaviour2 | SubjectB_behaviour2 | SubjectA_behaviour3 |

**Table 16. An example of a transition matrix, which is the result of a lag sequential analysis.**

Usually more than 5 states, as shown in ***Table 4*** will exist making the transition matrix larger and more difficult to overview and interpreted. We suggest a method for visualising the data in graph format. For this purpose we suggest to re-organise the matrix and divide it into 4 quadrants. Each of the quadrants gives the following information:

1. How many times does Subject A "response" to Subject B's behaviour. SSA initiated by Subject B.

2. How many times does Subject B perform one behaviour after another of Subject B's behaviour. Subject B's individual shifts between behaviours.

3. How many times does Subject A perform one behaviour after another of Subject A's behaviour. Subject A's individual shifts between behaviours

4. How many times does Subject B "response" to Subject A's behaviour. SSA initiated by Subject A

***Figure 20*** shows an example from the ATC study of this type of visualisation. (See also section 4.4)

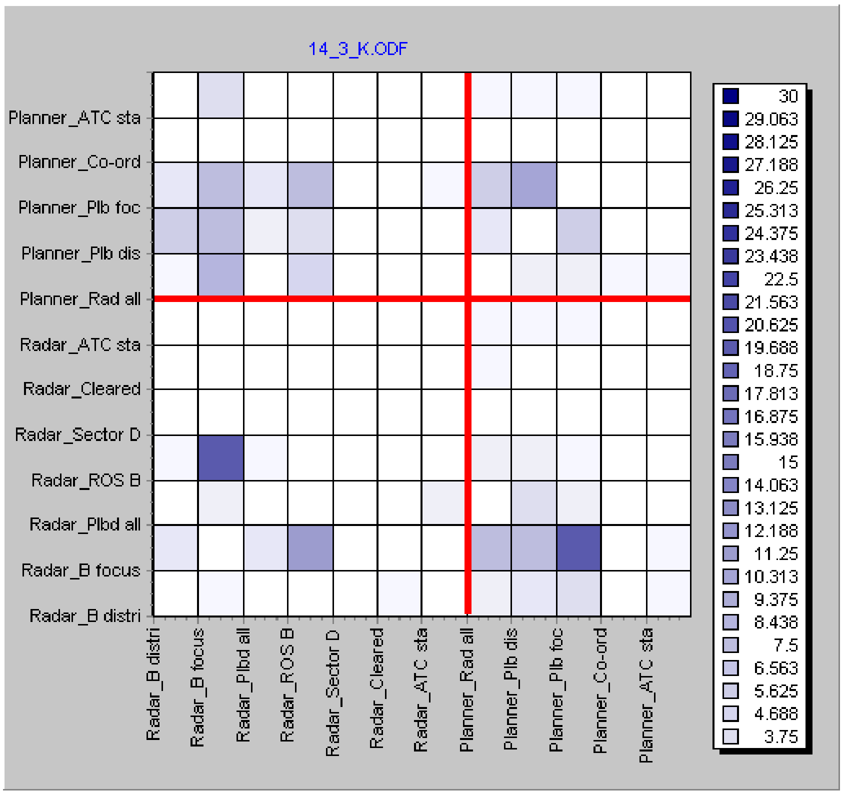



**Figure 20 Visualisation of SSA using lag sequential analysis.**

## 9.4 The Debriefing Questionnaire

We have devised a short battery of questions (ATSA: Assessment of Team Situation Awareness) designed to elicit a subjects' estimates of his or her team-mate's situation awareness and view of task allocation in addition to the subjects' own first-order knowledge of significant system parameters and their trends. The ATSA form *may* be used by interrupting a simulation session and eliciting subjects' estimates, continuing the session and then conducting a repeat interruption and elicitation of awareness judgements (like SAGAT was used). We suggest that it instead should be applied during post-session de-briefings.

When comparing responses from crews we shall be able to gauge

(d) the accuracy of individual crew members' estimates of system parameters (and, by extension, the agreement between crew members)

(e) the ability of crew members to correctly predict the awareness of their fellow crew member

(e) the extent to which crew members may correctly predict the workload of their colleague and their colleague's perception of task allocation.

The actual lay-out and content of the ATSA form can be found in Section 6.3.5. ***Table 17*** sums the measures that can be elicited in the assessment of team situation awareness by using the form.

**Table 17 TSA measures that can be elicited in the assessment of team situation awareness**

| **Assessment of Team Situation Awareness - table of measures elicited** |
|---|
| |

| **Measure** | **Individual SA measure:** Respondent's individual awareness | **Team Measure: Awareness of partner's awareness (TSA)** |
|---|---|---|
| workload | 5 point scale | 5 point scale |
| values of system parameters at time of interrupt | for each parameter, current | for each parameter, is my partner right / reasonably right / possibly far off |
| trend of system parameters (within last couple of minutes) | indicate whether parameter has been rising, falling or been stable | **nil** |
| your confidence in your own parameter estimate | 10 point scale | **nil** |
| who is responsible for monitoring this parameter during ***this*** phase | indicate whether responsibility for parameter monitoring is shared, respondent's or colleague's | indicate whether colleague will agree with task allocation (parameter monitoring responsibility) |

# 10 Acknowledgements

This work was partly funded by the European Community under the 'Competitive and Sustainable Growth' Programme (1998-2002). The following partners took part in the ***VINTHEC II*** consortium:
National Aerospace Laboratory NLR, BAE SYSTEMS, QinetiQ Ltd., National Defence Research Establishment, Linköping University, Risø National Laboratory, Technion – Israel Institute of Technology, Institute of Biocybernetics and Biomedical Engineering.

# 11 References

Adams, M., Tenney, Y., and Pew, R., Situation Awareness and the Cognitive Management of Complex Systems, *Human Factors*, 37(1), 85-104, 1995.

Andersen, H. (1997): Cooperative Documentation Production in Engineering Design: The 'Mechanisms of Interaction' Perspective, Topics in Cognitive Science and HCI, Centre for Cognitive Informatics. Roskilde University, (9).

Andersen, H.B. (2001) "Team Situation Awareness. Definitions and Measures. Technical Report R-1259, Risø National Laboratory, 4000 Roskilde Denmark. 2001.

Andersen, H.B. and Andersen, H. (2001): Team Situation Awareness as Mutual Knowledge. Risø National Laboratory, Roskilde. Risø-R-1206.

Andersen, H.B., Garde, H., Andersen, V. (1998). "MMS: An electronic message management system for emergency response". IEEE Trans. Eng. Manag., 45, 132-140.

Andersen, H.B., Sørensen, P.K., Weber, S. and Sørensen, C. (1996): A Study of the Performance of Captains and Crews in a Full Mission Simulator. Risø Technical Report (R-916). Risø National Laboratory, Roskilde, 4000 Denmark

Andersen, H.B., Sørensen, P.K., Weber, S., and Sørensen, C. (1996): A study of the performance of captains and crews in a full mission simulator. Risø National Laboratory, Roskilde. Risø-R-916.

Andersen, H.H.K. and Hauland, G. (2000): "Measuring Team Situation Awareness of Reactor Operators During Normal Operation: a Technical Pilot Study. In: Proceedings of the First Human Performance, Situation Awareness and Automation Conference, Georgia, Oct. 15-19, 268-273.

Bannon, Liam: "CSCW: An Initial Exploration," *Scandinavian Journal of Information Systems*, vol. 5, 1993, pp. 3-24

Bannon, Liam: "From Cognitive Science to Cooperative Design," *Theories and Technologies of the Knowledge Society, Centre for Cultural Research; Aarhus; Denmark*, edited by N. Finnemann, O., Centre for Cultural Research; Aarhus; Denmark, 1989c, pp. 33-58

Bannon, Liam: *Community Technology?: Issues in Computer-Supported Work*, Computer Science Department; Aarhus University; Aarhus; Denmark, 1989a.

Bannon, Liam: *Discovering the Human Actors in Human Factors*, Computer Science Department; Aarhus University; Aarhus; Denmark, 1989b.

Barresi,J. & Moore,C. (1996). Intentional relations and social understanding. Behavioral and Brain Sciences, 19(1), 107-122.

Borning, A., and Travers, M., Two Approaches to Casual Interaction over Computer and Video Networks, Proceedings of the Conference on Human Factors in Computing Systems, New Orleans, LA, 1991, 13-19.

Clark, H., Using Language, Cambridge University Press, Cambridge, 1996.

Clark, H.H. (1996). Using Language. Cambridge: Cambridge University Press.

Comstock, J.R. & Arnegard, R.J., 1992. *Multi-attribute task battery for human operator workload and strategic behaviour research.* (NASA Technical Memorandum 104174).Hampton, VA: NASA Langley Research Center.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J., (2000) Measuring Team Knowledge, In *Human Factors* VOL 42, No 1, Spring 2000 pp 151-173. Human Factors and Ergonomics Society.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., and Stout, R. J.(2000). "Measuring Team Knowledge". Human Factors, 42, (1), 151-173.

Costley, J., Johnson, D., and Lawson, D. (1989): A comparison of cockpit communication B737-B757. *Proceedings of the 5th International Symposium on Aviation Psychology* (pp 413-418). Columbus OH.

Dix, A., Finlay, J., Abowd, G., and Beale, R., Human-Computer Interaction, Prentice Hall, 1993.

Dourish, P., and Bly, S., Portholes: Supporting Awareness in a Distributed Work Group, Proceedings of the Conference on Human Factors in Computing Systems, Monterey, CA, 1992, 541-547.

Endsley, M. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems", Human Factors, 37, (1), 32-64,

Endsley, M. R., 1988. Situation awareness global assessment technique (SAGAT). In *Proceedings of the National Aerospace and Electronics Conference,* pp. 789-795. New York: IEEE.

Endsley, M. R., 2000. Theoretical Underpinnings of Situational Awareness: A Critical Review.In *Situation Awareness Analysis and Measurement.* M.R.Endsley & D.J.Garland (Eds.)Lawrence Erlbaum Associates, Inc., Publishers, Mahwah. Pp. 3-32

Endsley, M., Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors*, 37(1), 32-64, 1995.

Endsley, M.R. (1993): Situation Awareness in Dynamic Human Decision Making. Proceedings of the 1st International Conference on Situational Awareness in Complex Systems, Orlando, February 1993.

Endsley, M.R. (1995) Towards a Theory of Situation Awareness. Human Factors, 37 (1), 32-64

Flor, Nick V. and Hutchins, Edwin L. (1991). Analyzing Distributed Cognition in Software Teams: A Case Study of Team Programming During Perfective Software Maintenance. In Jurgen Koenemann-Belliveau, Thomas G. Moher, and Scott P. Robertson (eds.) Empirical Studies of Programmers: Fourth Workshop. Norwood, NJ: Ablex Publishing Corporation. pp. 36-63.

Foushee, H.C. , Lauber, J.K., Baetke, M.M., and Acomb, D.B. (1986): *Crew performance as function of exposure to high density, short-haul duty cycles.* (NASA Technical Memorandum 88322). Moffett Field, CA: NASA-Ames Research Center

Foushee, H.C. and Manos, K.L. (1981): Information transfer within the cockpit: Problems in intracockpit communications. In C.E. Billings & E.S. Cheaney (Eds.), *Information transfer problems in the aviation system* (NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center

Fracker, M.L., 1988. A theory of situation assessment: implications for measuring situation awareness. In *Proceedings of the Human Factors Society, 32nd Annual Meeting,* pp. 102-106. Santa Monica, CA: The Human Factors Society.

Fuchs, L., Pankoke-Babtz, U, & Prinz, W. (1995), Supporting Cooperative Awareness With Local Event Mechanisms: The GroupDesk System. In *Proceedings of the Fourth European Conference on Computer Supported Cooperative Work.* pp 247-262, Stockholm, Sweden

Gerson, Elihu M., and Susan Leigh Star: “Analyzing Due Process in the Workplace,” *TOIS*, vol. 4, no. 3, 1986, pp. 257-270.

Goodwin, C., Conversational Organization: Interaction Between Speakers and Hearers, Academic Press, New York, 1981

Greenberg, S., Peepholes: Low Cost Awareness of One’s Community, Proceedings of the Conference on Human Factors in Computing Systems (Conference Companion), Vancouver, 1996, 206-207

Greif, Irene (ed.): *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, San Mateo, California, 1988.

Grice, P. (1957). "Meaning". Philosophical Review, 66. 377-388.

Gutwin, C. and Greenberg, S. (1998) The Effects of Workspace Awareness Support on the Usability of Real-Time Distributed Groupware. Research Report 98-632-23, Computer Science Department, University of Calgary.

Gutwin, C. and Greenberg, S. (1999) A Framework of Awareness for Small Groups in Shared-Workspace Groupware. Research Report 99-2, Computer Science Department, University of Saskatchewan.

Hansen, J. P. (1991). "The Use of Eye Mark Recordings to Support Verbal Retrospection in Software Testing". Acta Pcychologica. 76, 31 – 49.

Harper, Richard R., John A. Hughes, and Dan Z. Shapiro. *The Functionality of Flight Strips in ATC Work. The Report for the Civil Aviation Authority*, Lancaster Sociotechnics Group, Department of Sociology, Lancaster University, January, 1989.

Harris, R. L., Sr., Glover, B. J., Spady, A. A., Jr., 1986. Analytical Techniques of Pilot Scanning Behavior and Their Application. NASA TP-2525.

Hauland, G. (1996): Building a Methodology for Studying Cognition in Process Control: a Semantic Analysis of Visual and Verbal Behaviour. Risø National Laboratory. Roskilde.

Heath, C. and Luff, P. (2000). Technology in Action. Cambridge: Cambridge University Press.

Heath, Christian, and Paul Luff: “Collaboration and Control. Crisis Management and Multimedia Technology in London Underground Control Rooms,” *Computer Supported Cooperative Work (CSCW). An International Journal,* vol. 1, no. 1-2, 1992

Heath, Christian, and Paul Luff: “Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms,” in *ECSCW ’91. Proceedings of the Second European Conference on Computer-Supported Cooperative Work*, ed. by L. Bannon, M. Robinson and K. Schmidt, Kluwer Academic Publishers, Amsterdam, 1991, pp. 65-80.

Helmreich, R.L and Foushee, H.C. (1993): Why crew resource management? In E.L. Wiener, BL.Kanki, & R.L Helmreich, R.L. (Eds) *Cockpit Resource Management*. Academic Press, New York.

Hilburn, B., 1996. *The impact of ATC decision aiding automation on controller workload and human-machine system performance.* NLR Tech. Pub. TP96999.

Holland, M.K., & Tarlow. G., 1972. Blinking and Mental Load. *Psychological Reports.* Vol. 31, pp. 119-127.

Hood, B.M., Willen, J.D. and Driver, J. (1998). "Adult Eyes Trigger Shifts of Visual Attention in Infants". Psych. Sci., 9, (2), 131-134

Hughes, John A., Dan Z. Shapiro, Wes W. Sharrock, and Robert Anderson: *The Automation of Air Traffic Control*, Lancaster Sociotechnics Group, Department of Sociology, Lancaster University, October, 1988.

Hutchins, E., The Technology of Team Navigation, in Intellectual Teamwork: Social and Technological Foundations of Cooperative Work, J. Galegher, R. Kraut and C. Egido ed., 191-220, Lawrence Erlbaum, Hillsdale, NJ, 1990.

Johansen, Robert: *Groupware. Computer Support for Business Teams,* The Free Press, New York & London, 1988

Jones, D. G., Endsley, M.R., 1996. Sources of situation awareness errors in aviation. *Aviation, Space, and Environmental Medicine, vol. 67,* pp. 507-512.

Kanki, B.G., Greaud, V.A., and Irwin, C.M. (1991): Communication variations and aircrew performance. *International Journal of Aviation Psychology 1,* 2, 149-162

Kanki, B.G., and Foushee, H.C. (1989): Communication as a group process mediator of aircrew performance. *Aviation, Space and Environmental Medicine, 60,* 402-410

Kanki, B.G., and Palmer, M.T. (1993): Communication and crew resource management. In E.L. Wiener, BL.Kanki, & R.L Helmreich, R.L. (Eds), *Cockpit Resource Management*. Academic Press, New York.

Kanki, B.G., Lozito, S.C., and Foushee, H.C. (1989): Communication indices of crew coordination. *Aviation, Space and Environmental Medicine, 60,* 56-60

Langton, S.H.R., Watt R.J. and Bruce, V. (2000). "Do the Eyes Have It? Cues to the Direction of Social Attention". Trends in Cognitive Sciences, 4, (2), 50-59.

Matthews, G., Middleton, W., Gilmartin, B., and Bullimore M. A., 1991. Pupillary diameter cognitive load. *Journal of Psychophysiology,* vol.5, pp. 265-271.

Middleton, D.: "Talking Work: Argument in Co-ordination and Improvisation In Team Work," *Summer Conference on Work and Communication, University of California; San Diego; California*, 1988.

Neisser, U., *Cognition and Reality*, W.H. Freeman, San Fransisco, 1976.

Norman, D., Things That Make Us Smart, Addison-Wesley, Reading, Mass., 1993.

Orasanu, J. (1990) *Shared mental models and crew decision making. Report No. 46.* Princeton NJ.

Oser, R.L., Prince, C., and Morgan, B.B. (1990): *Differences in aircrew communication content as a function of flight requirement: Implications for operational aircrew training*. Poster presented at the 34th annual meeting of the Human Factors Society, Orlando FL. [Cited from Prince et al., 1997]

Pedersen, C.R., Andersen, H.H.K. and Andersen, H.B. (2001) "Team Situation Awareness and Gaze". Technical Report R-1269, Risø National Laboratory, 4000 Roskilde Denmark. 2001

Pobertson, M. M. & Endsley, M. R., (1997,) Development of a situation awareness training program for aviation maintenance. In *Proceedings of The Human Factors and Ergonomics Society 41st Anual Meeting.* (pp 1163-1167). Santa Monica CA. Human Factors and Ergonomics Society.

Prince, A., Brannick, M.T., Prince, C. and Salas, E. (1997): The measurement of team process behaviours in the cockpit: Lessons learned. In M.T. Brannick, E. Salas & C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory, Methods and Applications.* L. Erlbaum Assoc. Mahwah NJ; New York.

Rasmussen, J., Pejtersen, A.M., & L.P., Goodstein (1994). Cognitive Systems Engineering. New York: John Wiley.

Rodgers, M. D., Mogford, R.H., Strauch, B., 2000. Post Hoc Assessment of Situation Awareness in Air Traffic Control Incidents and Major Aircraft Accidents. In *Situation Awareness Analysis and Measurement.* M.R.Endsley & D.J.Garland (Eds.) Lawrence Erlbaum Associates, Inc., Publishers, Mahwah. Pp. 73-112.

Ruffell-Smith, H.P. (1979): *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions.* (NASA Technical Memorandum 78482). Moffett Field, CA: NASA-Ames Research Center

Salas, E., Prince, C., Baker, D.P., Shrestha L., 1995. Situation Awareness in Team Performance: Implications for Measurement and training. *Human Factors, vol. 37,*pp. 123-136. (Cannon-Bowers and Salas, 1990)

Schmidt, Kjeld: *Social Mechanisms of Interaction*, Deliverable, Computing Department, Lancaster University, September, 1994b.

Segal, L., Effects of Checklist Interface on Non-Verbal Crew Communications, NASA Ames Research Center, Contractor Report 177639, 1994.

Strauss, Anselm, Shizuko Fagerhaugh, Barbara Suczek, and Carolyn Wiener: *Social Organization of Medical Work*, University of Chicago Press, Chicago and London, 1985.

Suchman, Lucy A.: "Office Procedures as Practical Action: Models of Work and System Design," *TOIS*, vol. 1, no. 4, 1983, pp. 320-328.

Suchman, Lucy A.: *Plans and situated actions. The problem of human-machine communication*, Cambridge University Press, Cambridge, 1987.

Tanaka, Y. & Yamaoka, K., 1993. Blink Activity and Task Difficulty. *Perceptual and Motor Skills.* Vol.77, pp. 55-66.

Veltman, J. A., Gaillard, A. W. K., 1998. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics,* vol. 41(5), pp. 656 -669.

VINTHEC Consortium, 1996. *Technical Report: Human Awareness Criteria*. VINTHEC-WP1-TR 01, NLR, Amsterdam.

VINTHEC Consortium, 1998. *Technical report: Pilot Studies*. VINTHEC-WP6-TR 01, NLR, Amsterdam.

VINTHEC II Consortium, 2001. *Technical report: State of the Art* VINTHEC II-WP1-TR 01, NLR, Amsterdam.

Wiener, E.L (1993): Crew coordination and training in advanced-technology cockpit. In E.L. Wiener, BL.Kanki, & R.L Helmreich, R.L. (Eds), *Cockpit Resource Management*. Academic Press, New York

Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (1993): *Cockpit Resource Management*. Academic Press, New York

Wilson, G.F. & Eggemeier, F.T., 1991. Psychophysiological assessment of workload in multi-task environments. In *Multiple-task performance.* D.L. Damos (Ed.) Taylor & Francis, London. Pp.337-360.

Wynn, Eleanor: "Office Conversation as an Information Medium," Ph.D. dissertation, University of California, Berkeley, CA, 1979

# 12 Abbreviations

| | |
|---|---|
| ACC | Area Control Centre, En Route ATC |
| AOI | Area Of Interest |
| APP | Approach |
| APRON | Ground Control |
| ASL | Applied Science Laboratories (EMT system) |
| ATC | Air Traffic Control |
| ATCO | Air Traffic Control Operator |
| ATM | Air Traffic Management |
| ATS | Air Traffic Services |
| ATSA | Assessment of Team Situation Awareness |
| CAA | Civil Aviation Authorities (SLV) |

| | |
|---|---|
| CAAA | Civil Aviation Authorities Academy (SLV-skolen) |
| CPH | Copenhagen |
| CSCW | Computer Supported Co-operative Work’ |
| DIA | Divisional Information Assistant (London Underground) |
| DR3 | Danish Reactor Number 3 |
| EMT | Eye Movement Tracking |
| FIR | Flight Information Region |
| FL | Flight Level |
| FPB | Flight Progress Board |
| ICAO | International Civil Aviation Organisation |
| IFR | Instrument Flight Rules |
| ILS | Instrument Landing System |
| IP | Information Processing |
| IR | Infra Red |
| ITV | Internal Television |
| LAI | Local ATS Instruction |
| MD | Minidisk |
| Observer | Noldus Observer Video Pro 3.0 (analysis system) |
| P | Planner Controller |
| POG | Point-Of-Gaze |
| R | Radar Controller |
| RED | Remote Eye Tracking Device |
| ROD | Rate Of Descent |
| SA | Situational Awareness |
| sSA | Shared Situational Awareness |
| SACRI | Situation Awareness Control Room Inventory |
| SAGAT | Situation Awareness Global Assessment Technique |
| SART | Situation Awareness Rating Technique |
| SLV | Statens Luftfarts Væsen (CAA) |
| SME | Subject Matter Expert |
| SMI | Sensimotoric Instrument (EMT, including Pholemus / head track) |
| SOB | Souls On Board |
| TA | Task Analysis |
| TRM | Team Resource Management training |
| TSA | Team Situation(al) Awareness |

| | |
|---|---|
| TWR | Tower |
| USB | Universal Serial Bus |
| VFR | Visual Flight Rules |
| VOR | Very high frequency (VHF) Omni directional Radio beacon |
| WL | Workload |

## Appendix A

### Order of runs

| | **RUN 1** | | **RUN 2** | | **RUN 3** | | **RUN 4** | |
|---|---|---|---|---|---|---|---|---|
| | **S / I** | **file** | **S / I** | **file** | **S / I** | **file** | **S / I** | **file** |
| **Team 1** (01,02) | S | 3a | I | 4a | S | 4b | I | 3b |
| **Team 2** (03,04) | I | 3c | S | 4c | I | 4a | S | 3a |
| **Team 3** (05,06) | I | 4b | S | 3b | I | 3c | S | 4c |
| **Team 4** (07,08) | S | 4a | I | 3a | S | 3b | I | 4b |
| **Team 5** (09.10) | S | 3c | I | 4c | S | 4a | I | 3a |
| **Team 6** (11,12) | I | 3b | S | 4b | I | 4c | S | 3c |
| **Team 7** (13,14) | I | 4a | S | 3a | I | 3b | S | 4b |
| **Team 8** (15,16) | S | 4c | I | 3c | S | 3a | I | 4a |
| **Team 9** (17,18) | S | 3b | I | 4b | S | 4c | I | 3c |
| **Team 10** (19,20) | I | 3a | S | 4a | I | 4b | S | 3b |
| **Team 11** (21,22) | I | 4c | S | 3c | I | 3a | S | 4a |
| **Team 12** (23,24) | S | 4b | I | 3b | S | 3c | I | 4c |

Note: S – co-operative ("samenwerking"), I – solitary ("individueel")

## Appendix B

### VINTHEC II  WP2 small scale experiment instructions to subjects

General introduction
The overall purpose of this experiment is to gather information about the feasibility of measuring a concept, which is called “shared Situational Awareness”, in the cockpit. The current experiment is a precursor of a larger scale experiment that will be performed in the NLR RFS. During this experiment you will work together with a colleague as a team. During the experimental runs each of you will work on different tasks, however the goal is that you, as a team, will perform optimal. Optimal performance in this experiment comes down to working fast and being alert, while making as few mistakes as possible.

Note that your responses to the computer-task as well as your scanning behavior will be recorded. The data will be anonimized and analyzed by NLR authorized staff only. The results will only be used for the current study.

The task that you will be performing is called the Multi Attribute Task (MAT) battery. It consists of four subtasks. Each window on the screen (see **Error! Reference source not found.**) presents a different subtask. Each subtask has some similarity with actual flight tasks. The two of you will focus primarily on different subtasks. The subtasks, which will be described in more detail below, are called: system monitoring, tracking, communication and resource management.

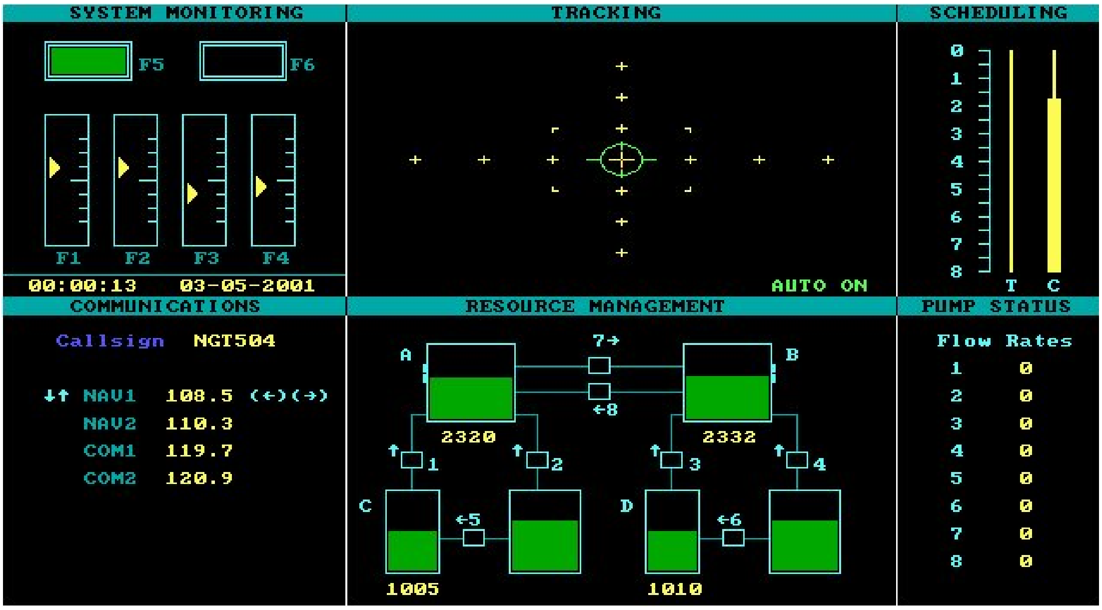



*Figure 21 The Multi Attribute Task (MAT) battery windows.*

Each of you will be responsible for the execution of two of those tasks. One of you (the co-pilot) will work on “system monitoring” and “tracking” while the other (the captain) will work on “communication” and “resource management”.

The tasks can be presented in either high- or low workload conditions. Sometimes the team will co-operate, and monitor each other, while in other settings each of you will focus on his / her own tasks. During the “co-operative” runs the captains’ task is to monitor, and if needed to assist, the co-pilot.

For each of the four different experimental runs you will receive specific instructions from the experiment leader.

**System monitoring (*co-pilot*)**
All of the information required to perform the monitoring task is displayed in the upper left window of the screen. This task consists of two parts: lights and dials. You will be monitoring the two lights at the top of this window for any changes. You will also be monitoring the four dials beneath them for any directional changes in the fluctuation of the pointer.

As you can see, during normal conditions, the left light is on in green. But occasionally this green light will go out. When this happens you must press the “F5” key as indicated next to that light. You will receive feedback in that the light will immediately turn back on.

The second light is normally off, but occasionally, a red light will turn on in this position. To respond to this, you must press the “F6” key, also indicated next to that light. As soon as you respond correctly, the red light will disappear.

The second part of this task consists of monitoring the four dials below the lights. Normally, the yellow pointers fluctuate form one unit below to one unit above the center line. Your task is to monitor these four dials and detect any change from the normal fluctuation of the pointer. In other words, if the pointer of one of these dials fluctuates either above or below the normal range, you must respond. The correct response is the key that is indicated below the dial which is out of range.

You’ll notice that feedback to a correct response is given by the presence of a yellow bar at the bottom of the dial that was out of range and a return to center of that dial pointer. Again, the abnormal fluctuation can occur in either direction - above or below - but there is only one response per dial.

**Tracking (*co-pilot*)**
All of the information that you need to perform the tracking task is displayed in the upper middle section of the screen in the section titled “Tracking”.

The overall purpose of this task is to keep the airplane symbol, represented by the green circle, within the dotted rectangular area in the center of this task.

If you do not control the plane with the mouse, the plane will drift away from the center. You must control the plane with movements of the mouse. Basically, you must compensate for this random drifting by pulling the plane back to center with corresponding movements with the mouse. For example, if the plane is drifting to the right, moving the mouse to the left will return the ship to center. You’ll notice that if the plane is away from the center, you must make rather large movements to return it. If the plane is already in the center, smaller movements will be required.

**Resource Management (*captain*)**
All of the information that you will need to do the resource management task is contained within the two lower right windows with the headings “Resource Management” and “Pump Status”.

This task is considered a fuel management task. The rectangles are tanks which hold fuel, the green levels within the tanks that increase and decrease are fuel, and along the lines which connect the tanks are pumps which transfer fuel from one tank to another in the direction that is indicated by the arrow. The numbers underneath four of the tanks represent the amount of fuel in units of each of these tanks. The number will be increasing and decreasing as these levels change. The capacity for the main tanks, A and B, is 4000 units each. The supply tanks on the right of each three-tank system have an unlimited capacity - they never run out.

Your overall goal with this task is to maintain the levels of fuel in tank A and B at 2500 units. This critical level is indicated by the thick mark in the shaded area on the side of each of these tanks. This level is also indicated by the numbers underneath each tank. It is acceptable to keep the level of fuel within the shaded area between 2000 and 3000 units. However, optimum performance is obtained when Tanks A and B are at 2500 units.

In order to meet this criterion, you must transfer fuel to tanks A and B in order to meet this criteria because tanks A and B lose fuel at the rate of 800 units per minute. So you can see that with their present levels of approximately 2400 units each, these tanks would become empty in slightly more than 3 minutes without the transfer of additional fuel. Tanks C and D only lose fuel if they are transferring fuel to another tank.

The process of transferring fuel is as follows. Notice that every pump has a number, a square box and an arrow next to it. The arrow indicates the direction through which fuel can be transferred with that pump. Each pump can only transfer fuel in one direction. The pumps are activated by pressing the key corresponding to the pump that you wish to activate. Use the number keys across the top of the keyboard rather than those on the right hand of the keyboard.

When the pumps are turned on, two things occur. First, the square of each pump turned green. That means that the pump is actively transferring fuel. When the pump is off, the square is black. The second change on the screen is the numbers that appeared in the "Pump Status" window. Let's focus on that now.

Under "Pump Status", two columns of numbers are presented. The first column, numbers one through eighth, indicate the pump numbers and these correspond directly to the pumps in the diagram. The second column of numbers indicates the flow rates in units per minute of each pump when that pump is on. For example, Pump 1 transfers 800 units of fuel per minute from Tank C to Tank A. The flow rate for any given pump is only presented if that pump is on and actively transferring fuel.

So far, you've seen two conditions for the pumps: on and off. Pressing the pump number key once turns the pump on; pressing the key again turns the pump off, and so on. A third condition is the fault condition, over which you have no control. At various times throughout your task, you'll see the status indicator on a pump turn red. This means the pump is inactive as long as that red light is present. You will not be able to use this pump until the red light goes out. However, you must be aware that when the fault is corrected and the red light goes out, that pump will be automatically returned to the "off" status (without any light). Even if you had turned that pump on before the fault occurred, the pump will not be returned to an "on" condition. You will have to turn it on again if that is what you wish.

Along the same line, if a tank fills up to its capacity, all incoming pump lines will be turned off automatically. This is because a full tank cannot receive any more fuel. You will have to turn those pumps back on at a later time, if that is what you wish. Conversely, if a tank becomes empty, all outgoing pumps will automatically turned off. This is because an empty tank can no longer transfer fuel. Again, you will have to turn these pumps on again if that is what you wish to do.

Your overall goal is to keep the fuel level in the Tanks A and B at 2500 units each. You may use any strategy that you wish to do so. If the fuel level in these tanks should go outside the shaded region, however, please return the fuel back to the target level as soon as possible.

**Scheduling Window (*both*)**

The scheduling window is found in the upper right corner of the display. The purpose of the scheduling window is to present the start and duration of the manual tracking task and the communication task. The scheduling window requires no response on your part but is designed to provide you with information about the scheduling of tasks in the near future. The two indicators are identified by “T” for the tracking task, and “C” for the communication task. The scheduling window allows you to “look ahead” from the present (0 minutes) to 8 minutes into the future. The bold lines, or bars, indicate times during which these two tasks, tracking and communication, will be operating. The thin lines indicate times at which either tracking or communication are not required.

Thank you for your co-operation and good luck!!!

## Mission

To promote an innovative and environmentally sustainable technological development within the areas of energy, industrial technology and bioproduction through research, innovation and advisory services.

## Vision

Risø's research **shall extend the boundaries** for the understanding of nature's processes and interactions right down to the molecular nanoscale.

The results obtained shall **set new trends** for the development of sustainable technologies within the fields of energy, industrial technology and biotechnology.

The efforts made **shall benefit** Danish society and lead to the development of new multi-billion industries.